NASA Technical Memorandum 89403

Space Research Data Management in the National Aeronautics and Space Administration

George H. Ludwig

NOVEMBER 1986

NASA

NASA Technical Memorandum 89403

Space Research Data Management in the National Aeronautics and Space Administration

George H. Ludwig
National Aeronautics and Space Administration
Washington, D.C.

National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1986

FOREWORD

The Committee on Data Management and Computation (CODMAC) of the Space Science Board of the National Research Council was founded in 1978 to advise the National Aeronautics and Space Administration's (NASA) Office of Space Science and Applications on ways to improve the handling of space science data and associated computational resources from the user point of view. CODMAC has, to data, published two reports (<u>Data management and Computation, Volume 1: Issues and Recommendations</u>, NAS, 1982; <u>Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences</u>, NAS, in press) which have described principles for data management, suggested guidelines for the operation of data repositories and centers, and made general policy and technology development recommendations. As part of its continuing examination of NASA data and information systems activities, the Committee has reviewed this document.

The CODMAC, as well as other groups, have urged that changes be made in the manner in which NASA manages its space science data activities. George Ludwig's report represents an internal NASA study which attempts to identify specific steps which NASA might take to improve space science and applications data management in response to these external reviews. As such it is consistent with the contents of the two CODMAC reports. It goes a step beyond our reports to make specific recommendations within NASA for improving the effectiveness of data system activities. These recommendations have been discussed with NASA data managers; they represent a set of actions that can be achieved.

The implementation of these recommendations would be a major step forward in improving the utilization of current space research data, and in preparing for many of the exciting challenges of the future. The Committee on Data Management and Computation of the National Academy of Sciences strongly endorses this report.

Christopher T. Russell
Chairman
Committee on Data Management
and Computation

EXECUTIVE SUMMARY

The National Aeronautics and Space Administration has established an enviable record of achievement in building, launching, and operating spacecraft for space science research. The results from those missions are known throughout the world. The extraction of those results required the establishment and use of complex technical and management arrangements for operating the spacecraft and for acquiring and processing their data.

Space-related scientific research has passed through a natural evolutionary process. While the opportunities for the unexpected basic discovery and the need for relatively simple experiments to map newly discovered phenomena remain, most space research today is much more complex. Modern efforts tend to involve the use of relatively large and elaborate instruments for studies of broad scope. The operation of some of those instruments involves real time or near real time decision-making by the investigators. Many of the lines of investigation require the collaborative use of data from multiple sensors by widely dispersed research teams. The task of extracting the meaningful information from the raw data is highly involved. Looking into the future, many of the space research endeavors will require data processing capabilities which simply do not exist today.

In recent years there have been numerous studies of NASA's practices in managing space research data. This present document reports the results of a three-year examination of this subject, using the earlier reports as a starting point. The general conclusions are that (1) there are areas in which improvements in NASA's space research data management practices would enhance the ability to extract more of the potential value from the space-related data, and (2) the agency's ability to substantially capitalize on some of the exciting opportunities of the future critically depend on such changes.

The report text describes the envisioned future research environment and the need for new approaches for operating in that environment. It contains a number of observations about the current state of affairs, and recommends specific NASA actions. Those observations and recommendations, as extracted below, provide a comprehensive overview of the report.

OBSERVATIONS:

1. During more than 28 years of space research, many lines of investigation have advanced to the point where further progress depends on approaches which (1) depend on increasingly discriminating observations, (2) involve expanded physical scales, (3) encompass extended periods of time, and (4) are disciplinary or multidisciplinary in nature, requiring the merging of data from multiple sources. There is broad agreement that new technical means for handling space and earth sciences missions data are needed to meet these needs.

2. Each of the investigator models (principal investigator, investigator team, guest investigator, researcher consortium, and retrospective user) contains features which serve particular needs. Each will persist into the future, including the space station era. The information systems should continue to evolve to serve this variety of research environments.

3. There is increasing agreement that space research data management planning should be treated as a system problem, involving the concept of distributed Space Science Data Management Units with a hierarchy of databases and integrating networks.

4. Data format, system interface, and documentation guidelines and standards are urgently needed for directories, catalogs, and their inquiry networks to assist users in acquiring information about data. The development of guidelines and standards for the data repositories and archives is also desirable. NASA communication circuit format standards are in existence or being developed; comparable guidelines and standards are needed for the data interfaces between many other elements of the space research data systems. The adoption of guidelines and standards for some of the software associated with instrument checkout and operation, data distribution, and databases should be critically examined. Selected documentation guidelines and standards are also needed.

5. The lack of appropriate basic technologies does not appear to be a major factor limiting present space research data use. The diligent application of available technologies, coupled with suitable planning, coordination, and management, could satisfy most of the needs for the presently approved missions. Carefully selected technology advancement is needed, however, for some of the longer-range future missions, including some of those envisioned for the space station. Areas needing attention include (1) on-board data processing, (2) distributed interactive networks for science operations and analysis, (3) very large mass data storage, (4) database structures, including the use of database machines, (5) transportable software, and (6) the application of artificial intelligence approaches. NASA needs to develop and maintain a keen awareness of technology developments in these areas. In selected instances, NASA should fund technology development which is needed for its future missions, and which would not otherwise be forthcoming.

6. The present National Aeronautics and Space Administration organizational structure, in general, serves the agency well. Although the space research data management responsibilities are widely dispersed, there are no fundamental reasons why they cannot be adequately coordinated to provide cohesive, well-integrated data systems to meet present and future needs. It is suggested that only minor changes to the present overall NASA organizational structure are needed for the purpose of improving space research data management.

7. NASA has excelled in planning and managing the development of spacecraft and the directly supporting ground subsystems required for their initial activation. An amazing amount of new scientific information has been gleaned from its space research missions. A majority of the specific

science objectives for the individual missions have been met. It may have been possible to achieve even more of the mission potential through better planning and execution of the data management functions.

8. NASA has taken many specific steps to make improved use of space research data and to facilitate the more complex missions and the broader classes of study presently envisioned. But an overall, cohesive, long-term vision of this data management task, including a realistic understanding at the highest organizational levels of the resources and management attention required, has been lacking. The absence of cohesive long-range planning has been a persistent aspect of NASA's space research data management.

9. To be effective, space research data management systems must be planned, built, and operated through an active partnership between the technical professionals and the researchers. NASA has an obligation to provide opportunities for meaningful researcher participation in planning and building the systems. In turn, the researchers have an obligation to devote the time and energy required to actively assist.

RECOMMENDATIONS:

Related to Planning and Management:

1. It is recommended that the Office of Space Science and Applications undertake, as a high priority task, the development of a long-range strategic plan for space research data management. This plan should encompass all phases of data handling and analysis from instrument checkout and mission operation to data archiving, and span the entire system from the sensors through the databases. It should concentrate on general principles and capabilities, rather than detailed designs. Preparation of the plan should actively involve the headquarters discipline directors, program managers, field center project personnel, information system professionals, and external research communities. The plan should embody the system concepts and the principles for successful scientific data management outlined in Section 2.4 of the report text. Once approved, this plan should serve as the basis for specific mission and discipline data system planning and budget formulation.

2. It is recommended that the Associate Administrator for Space Science and Applications assign to designated officials specific responsibilities (with accompanying resources, including field center support) in the areas of data management long-range strategic planning, data policy, guidelines and standards, data system budgeting, oversight, advanced development, and performance evaluation, as outlined in Section 6 of the report text.

3. It is recommended that an external peer review group be charged with annually (1) reviewing NASA's progress in information system planning and implementation, (2) evaluating system performance, (3) assessing conformance with data policy directives, and (4) suggesting changes and new directions for the future. The review group membership should be roughly balanced between practicing space researchers and information system professionals.

Related to Policy Guidance:

1. It is recommended that the policy elements of NASA Management Instruction 8030.3a, expanded as necessary, be incorporated in a short NASA Policy Directive. The remaining portions of the NMI, including the background, explanatory materials, implementation instructions, and delegation of responsibilities, should be updated to reflect the changes which have occurred, and which are envisioned for the future.

2. It is recommended that the Office of Space Science and Applications take specific steps to enforce the provisions of the resulting policy and implementation guidance.

Related to Data Centers:

1. It is recommended that the Office of Space Science and Applications, with participation by the field centers and external researchers, develop a cohesive plan for the data center activities, including not only the National Space Science Data Center, but the other data holdings for which long-term retention and accessibility is desired. These entities should be operated as a dynamic, integrated network in a manner which will cause them to play a more active role in the research process. One of the most urgent first steps should be to work out a cohesive, viable funding structure for all data center activities.

2. The NSSDC, in addition to its other functions, should be responsible for (1) establishing and operating a central data directory and inquiry system, (2) leading the establishment of catalog, archive, and documentation guidelines and standards for use by all of the NASA data centers, and (3) arranging for the exchange of catalogs and data with other agencies and countries.

Related to Standards:

It is recommended that the Office of Space Science and Applications, working closely with the NASA field centers and external research communities, prepare a plan for the evolutionary development and promulgation of carefully selected data format, system interface, software, and documentation guidelines and standards. This plan should encompass selected major space science data system interfaces, repositories, archives, data directories, catalogs, and their documentation. It should include the establishment of libraries for exchanging frequently used investigators' utility software. Responsibilities should be clearly assigned for developing and carrying out this plan.

ACKNOWLEDGMENTS

This study of the National Aeronautics and Space Administration space research data management practices and policies was conducted over a period of time extending from June 1983 through July 1986. One of the reasons that it took such a long time was to permit extensive discussions with the management officials, researchers, and information system professionals who are all participants in the use of space-derived data. It was hoped that this interaction might result in a report which would have a wide base of acceptance and support at the time of its publication.

Many people took part in those discussions. Their encouragement and help is acknowledged with grateful thanks.

Frank B. McDonald, NASA Chief Scientist, initially encouraged me to undertake this study and provided support for it throughout the three year period. His wise counsel and continuing support are immensely appreciated.

During the initial eight month period, while I served as a member of Dr. McDonald's staff and was physically located in the Washington area, I was fortunate to have had the assistance of a very capable Study Team, assembled for this purpose. The members provided general advice and specific comments on early drafts of the report. They are, in alphabetical order:

Alex Dula	NASA Headquarters
John W. Dyer	Ames Research Center
Joseph Fuller, Jr.	Goddard Space Flight Center
Joseph H. King	Goddard Space Flight Center
Donald A. Krueger	Goddard Space Flight Center
James E. Kupperian, Jr.	Goddard Space Flight Center
Caldwell McCoy, Jr.	NASA Headquarters
William H. Mish	Goddard Space Flight Center
George F. Pieper	Goddard Space Flight Center
John Y. Sos	Goddard Space Flight Center
Charles H. Stembridge	Jet Propulsion Laboratory
Robert R. Stephens	NASA Headquarters
Michael J. Wiskerchen	NASA Headquarters

A number of other individuals who helped by providing written comments on draft copies of the report are:

Raymond E. Arvidson	Garry Hunt
Roger G. Barry	J. Charles Klose
Fred C. Billingsley	Lucy McFadden
Lawrence Bolef	Christopher Russell
Robert R. P. Chase	Ethan Schreier
Michael W. Devirian	James I. Vette
John E. Estes	Gio Weiderhold
Ida R. Hoos	James Weiss

Contact was maintained with several key groups active in space research data management. Progress reports were made to them from time to time, and a number of their members provided comments. Without attempting to name their memberships at the times of the presentations, these groups are:

- Space Science Board, National Academy of Sciences/National Research Council
- Committee on Data Management and Computation, Space Science Board
- Space and Earth Sciences Advisory Committee, NASA Advisory Council
- Information Subcommittee, Space Applications Advisory Committee, NASA Advisory Council

Informative discussions were held with many other individuals on numerous occasions. Those contributors (with sincere apologies to the ones who may have been omitted) are:

T. Albert	R. C. Hart	G. F. Pieper
J. K. Alexander	W. N. Hess	W. L. Quaide
K. A. Anderson	N. W. Hinners	I. Rasool
R. J. Arnold	J. D. Hodge	J. D. Rosendhal
D. Atlas	S. S. Holt	S. Ruttenberg
R. C. Baumann	R. L. Jenne	P. H. Smith
J. H. Boeckel	S. W. Keller	R. E. Smylie
P. A. Bracken	J. H. King	C. W. Snyder
J. H. Bredekamp	J. E. Kupperian, Jr.	S. Sobieski
F. P. Bretherton	D. A. Landgrebe	J. Y. Sos
G. A. Briggs	A. L. Lane	R. R. Stephens
M. A. Chinnery	H. R. Lynn	E. C. Stone
V. E. Derr	H. Mark	V. E. Suomi
T. M. Donahue	F. D. Martin	J. S. Theon
A. Dula	T. B. McCord	S. G. Tilford
J. H. Duxbury	C. McCoy, Jr.	J. H. Trainor
B. I. Edleson	C. E. McIlwain	J. A. Van Allen
K. J. Frost	L. H. Meredith	A. J. Villasenor
R. Giacconi	N. F. Ness	R. J. Walker
A. K. Guha	M. M. Neugebauer	G. Wiederhold
T. O. Haig	C. J. Pellerin	M. J. Wiskerchen

TABLE OF CONTENTS

1. INTRODUCTION

This document reports the results of an examination of the National Aeronautics and Space Administration's (NASA's) technical and management practices and policies for obtaining, distributing, processing, storing, and otherwise using space-derived and associated data for research purposes. This study was conducted during the period from June 1983 through July 1986.

The study goal was to develop an understanding of the changing character of space research, of NASA's practices in managing space research data, and to make recommendations which might improve the effectiveness of future space research through increased data utilization.

It is acknowledged that other federal agencies carry related responsibilities for space research data management. An especially important case is the operation of a suite of environmental data centers by the National Oceanic and Atmospheric Administration. It is assumed throughout this report that the NASA data management policies and practices will continue to be in accord with the memorandums of understanding and other documents which define such collaborative relationships.

1.1. Background

Spectacular progress has been achieved during the past twenty-eight years in the use of space-based sensors. Automated probes have observed most of the solar system's planets. Landed instruments have made surface observations on the Earth's moon and Mars. Earth orbiting spacecraft have measured a diversity of phenomena ranging from the radiation reaching our neighborhood from the far reaches of the universe, to the complex interplay between charged particles and magnetic fields in the solar system, to the earth's weather with its associated surface interactions. Land-observing instruments have given us a new way to detect and quantify many features on the earth's surface.

Analyses of those observations and measurements has led to a much better understanding of our physical universe and the processes which shape it. Future substantial advancement along these lines depends in a most crucial manner upon continuing improvements in our capabilities for remote and *in situ* sensing, and for manipulating and interpreting the resultant data. Building the space station will provide an added stimulation for the development of new classes of observational and data management capabilities.

In spite of this impressive record of achievement, there have been persistent concerns that we may not have been realizing the full potential of the space research data. These concerns have pertained to (1) specific technical arrangements and timeliness for providing mission data to the researchers, (2) easy access by the researchers to correlative and ancillary data, (3) arrangements for data storage and archiving, and (4) inadequate funding for data reduction and analysis. Some investigators fear that new classes of investigations presently envisioned may be difficult or impossible for lack of coherent data system planning and purposeful execution of those plans.

These concerns have been the subject of numerous studies conducted both within and outside the agency. Key documents which have been drawn upon heavily in this present review are two reports prepared by the National Academy of Science's Committee on Data Management and Computation (CODMAC). The first is entitled Data Management and Computation: Issues and Recommendations [NRC, 1982], while the second is labeled Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences [NRC, 1986]. These, along with the other published reports which are cited in the text, are listed in the reference section. Appendix A contains summaries of additional reports which influenced this study, but are not in the easily available literature. In the aggregate, these reports provide a wealth of information, including statements of issues, related discussions, and recommendations for NASA action. Although some of the individual issues and recommendations might be argued, when viewed as a whole they provide meaningful indications of areas which deserve added attention, including promising opportunities for the future which should not be ignored.

Both in these reports and elsewhere, many researchers and managers have expressed their belief that NASA has been slow to move ahead in space research data management, and have cited examples of missions in which data management has been poor. In the interest of presenting a balanced picture, it is noted that a number of space research missions, especially in the recent past, have successfully applied modern principles of data management. Examples of such missions include the following:

- For the Solar Mesosphere Explorer (SME), Solar Maximum Mission (SMM), and Active Magnetosphere Particle Tracer Explorer (AMPTE) the mission control and major portions of the data processing have been handled directly by the principal investigators. They have had a great deal of success in meeting the unique scientific objectives of each of the missions.

- International Ultraviolet Explorer (IUE) operations and data analysis have been remarkably successful, and have been highlighted on practically every visit by visiting dignitaries to Goddard. The system is continuing to evolve so that observers can now conduct observations from their home institutions. IUE has made substantial progress in data archiving.

- High-Energy Astrophysical Observatories (HEAO) I and II were run under the control of the principal investigators, and the system was sufficiently responsive that new operational modes were implemented by the PI's which had not been envisaged by the spacecraft designers. HEAO data analysis and the guest observer program are still active almost five years after the last spacecraft ceased operating.

- In the Voyager I and II missions the investigators have been involved in all phases of planning for the planetary encounters and in establishing detailed timelines. There is general agreement among the investigators that their needs have been met.

- The International Sun-Earth Explorers (ISEE) 1, 2, and 3, Pioneers 10 and 11, Pioneer-Venus, and the Interplanetary Monitoring Platform (IMP) 8 all represent current missions where data are flowing in a timely manner, and where support for data analysis is available long after the period which was originally planned.

The primary purpose of this report is to highlight areas where further actions should place us in a position to take more complete advantage of the opportunities of the future.

1.2. A General Perspective

Stated in the broadest terms, one of NASA's primary functions is the conduct of an effective space and earth sciences research program which expands human knowledge about the earth, its environs, the solar system, and the universe as a whole. To achieve success the Agency must deal in a balanced manner with the entire research process, including (1) identifying problems for study, (2) planning specific programs and projects for attacking those problems, (3) developing, deploying, and using suitable instruments and facilities for gathering and analyzing the data, (4) making the data available, and (5) stimulating and facilitating the intellectual process of examining the data, building hypotheses, and critically examining them within the world research communities.

In this context, the data produced by space research projects represent a valuable and often unique national resource, acquired with the substantial expenditure of intellectual and fiscal resources. The analysis of these data is a complex and lengthy process, requiring the further commitment of substantial resources in order to realize a full return from the original expenditures. Many of the ultimate uses of the data cannot be foreseen when the programs are planned; frequently new ideas emerge long after the data are acquired, as a result of the continuously improving understanding of the physical processes under study. Even though some of the acquired data may never be fully used, it is usually not possible to decide in advance which data will ultimately be least or most important. Careful thought and planning by the researchers and managers is necessary to ensure that the full benefits, both short and long-term, can be realized in the face of these uncertainties. The need for this planning and for aggressive action to implement the resulting plans is becoming more and more pressing as the experiments become more complex, as the data volumes increase, and as we must take advantage of increasingly sophisticated technical capabilities in order to achieve our goals.

It is this complete process, with its emphasis on the ultimate research results, which should be NASA's primary focus as it strives to advance mankind's understanding of the physical universe. One of the primary objectives of this paper is to provide a convincing argument that the planning, building, and use of the information system is just as much a part of a mission as is the development, launch, and operation of a spacecraft with its instruments, and deserves just as much attention.

2. SPACE RESEARCH AND TECHNOLOGY IN A STATE OF CHANGE

Research progresses through stages as it matures. These are identified for present purposes as discovery, exploration, interpretation, and consolidation. To illustrate the process, the earth's Van Allen Radiation Belts were discovered in 1958 by the analysis of data from the cosmic ray detectors on Explorers I and III. The early exploration of those belts, resulting in their initial spatial delineation, the identification of the particle types, and the determination of the particle energy spectra, was conducted by the use of instruments on the subsequent Explorer spacecraft and the Orbiting Geophysical Observatories. These initial processes were observation-intensive. By the mid-1960's a sizable body of descriptive information had been published, and interpretive hypotheses dealing with source, injection, trapping, and decay mechanisms were evolving. The researchers came to realize that the trapped radiation could not be treated in isolation, and a consolidation of related fields began. Earlier work in geomagnetism and upper atmospheric physics was drawn upon, and as the profound effects within the magnetosphere of processes in the sun and the extended medium from the sun to beyond one astronomical unit became better understood, the broader field of solar-terrestrial physics emerged. It is now widely accepted that further progress in our understanding requires examination on the scale of the entire inner solar system. This evolutionary process is a natural feature of every branch of space research.

As the character and complexity of the experiments evolve, so do the demands on the related information systems. Some of the earliest experiments employed very simple on-board processing, low bandwidth dual frequency modulated telemetry, the manual reading of strip chart records, and analyses using hand calculators. Today, microcomputers are used in many space instruments, data are transmitted at rates of millions of binary digits per second, and elaborate ground processing programs assist in optimizing the instrument operation and in preparing the data for human examination. Thus, enabling technologies have emerged hand in hand with Space Science's evolution.

2.1. The Maturing of Space Research

The current and planned space and earth sciences programs include projects at each of the four previously described stages of development. As work progresses on older, more established lines of investigation, new discoveries are made which, in turn, stimulate new exploration, interpretation and consolidation. And from time to time substantial new capabilities, such as larger launch vehicles, appear on the scene to open entirely new lines of investigation.

A number of recent space and earth sciences planning documents reveal the extent to which this process has progressed in some disciplines, and the degree to which its occurrence is recognized by the associated scientists and program managers. For example, in the National Research Council's report Space Plasma Physics: The Study of Solar-System Plasmas, [NRC, 1978, p. 3] one of the principal conclusions is:

> "Now that the initial exploratory stage of space plasma physics has been completed successfully, the fruitfulness of future projects will depend on addressing basic scientific problems. The solution to these problems will call for a logical cycle of theoretical problems definition, the planning of experiments and hence missions, data collection, data reduction, and theoretical analysis, leading to a progressive refinement of the science."

The need for broadening the thrust of the research for the study of processes on and near the earth's surface is clearly acknowledged in the NASA report, The Pilot Land Data System: Report of the Program Planning Workshop [NASA, 1984a, p. vi]:

> "There is a trend in scientific research in general, and more specifically in National Aeronautics and Space Administration programs, to ask research questions which are multidisciplinary in nature and global in scale. Researchers at agencies and institutions across the

> nation and around the globe are attempting to improve our understanding of global carbon cycling; the relationship between land energy balance and biophysical conditions, and their interrelationship with climate; global and regional geologic and geomorphic structure and process; and, to identify early indications of change in global elemental cycles, climate, hydrology and environmental quality."

And in the Earth Observing System Science and Mission Requirements Working Group Report [NASA, 1984b, p. 1]:

> "Earth science is a maturing research field where much is known about a variety of detailed processes, but where the broad integrating themes and understanding are just now beginning to take shape. The past two decades have seen major progress with the concepts of dynamic meteorology and plate tectonics. Both developments required new observational techniques. With comparable improvements in measurement capabilities, breakthroughs appear possible in the understanding of ocean circulation and upper atmospheric processes during the 1980's. The decade of the 1990's will provide opportunities for similar progress on scientific problems that require an integrated approach to the study of Earth. This decade will also provide the opportunity to obtain simultaneous measurements of parameters relevant to energy disposition, hydrologic status, and chemical cycles on a global basis. Key examples of research areas which will benefit from this approach are the role of the polar regions in climate, the large scale hydrologic cycle, global biology, and biogeochemistry."

OBSERVATION: During more than 28 years of space research many lines of investigation have advanced to the point where further progress depends on approaches which (1) depend on increasingly discriminating observations, (2) involve expanded physical scales, (3) encompass long periods of time, requiring the accumulation of carefully calibrated and safeguarded data in order to distinguish long-term trends, and (4) are disciplinary or multidisciplinary in nature, requiring the merging of data from multiple sources.

2.2. Advances in Research Data Management

The implications of this changing character of space research are profound. It requires that the data sets be used in an increasingly complex and interrelated fashion. In some cases, just the control of the instrument operation and of the data acquisition process require sophisticated data processing, presentation, and response capabilities. It is no longer possible to plan the operation and data handling for individual sensors without, at the same time, planning the manner in which the data from a diversity of sources will be acquired, distributed, merged, intercalibrated, stored, and otherwise handled.

Here, too, a broadening awareness of this need has developed. The report Solar-Terrestrial Data Access, Distribution, and Archiving by the National Academy of Science's Joint Data Panel [NRC, 1984, p. 2] states:

> ". . . Since the discovery of the radiation belts, much successful space research has been accomplished by independently analyzing data from individual satellite-borne instruments. Great advances in our knowledge of the space environment were made using this approach, particularly during the discovery and exploration phases. Based on these advances, it became clear that the solutions to current scientific problems in solar-terrestrial research require data from a variety of instruments - a multiparameter data base - and individual researchers began to share data to pursue these problems jointly. The concept of obtaining a multiparameter data set suitable for particular space physics problems was fundamental in shaping the instrument complements and/or data handling systems in the Explorer 45 (SSS-A), AE, SMM, SME, and DE programs."

In planetary research a similar conclusion has been reached. The report, Trends in Planetary Data Analysis - Executive Summary of the Planetary Data Workshop, Nov. 29-Dec. 1, 1983 [NASA, 1984c] states:

> "The emergence of the Interdisciplinary Scientists (mission science members whose research responsibilities include the output from many instruments on the spacecraft) and the discovery of elements in the solar system too complex to be understood by isolated analysis attempts - Saturn's Rings, the Venus environment - have created new requirements for effective data analysis by the planetary community."

This reference strongly asserts that past approaches for data management are inadequate to meet the new needs:

> "The desire to compare one data set with another, at the highest possible resolution; to view in synopsis, temporal or spatial, all observations of an object or phenomena; to transfer raw data bits or analyzed results rapidly between investigators; and to coordinate results and reexamine details in real time; place requirements upon the data set and the current data analysis system that cannot be met. The

(present) archaic system cannot support modern data analysis techniques. . . ."

The National Academy of Science's Committee on Data Management and Computation, in their second report [NRC, 1986] observes:

> "Research within the disciplines covered by the space sciences is moving into a new era, one in which a large volume of data will be acquired and where a number of data sets will be needed within a variety of . . . settings to solve the increasingly complex questions that are being addressed. Generally, the data volumes and data uses will grow much faster than the number of researchers examining the data. In addition, the data needed for any given task may not have been collected by any one researcher. Thus, ready access to high quality, well documented data, together with the ability to handle, process, and store the data, are key ingredients for successful management of space sciences data in the 1980's and 1990's. Advances in computation and communications will allow such data management to be done in new, innovative ways."

The present availability of the Space Transportation System and the planned development of the space station and tended platforms open exciting new possibilities for placing or building very large instruments in orbit, resupplying and maintaining them to assure longevity, and operating them to optimize their productivity. These provide an additional impetus for developing new ways of building and using the supporting information systems.

OBSERVATION: **There is broad agreement that new technical means for handling the space and earth sciences missions data are needed to meet these needs.**

2.3. Investigator Models

Investigators interact with the space research process in several ways. The <u>principal investigator</u> (PI) model employed for many experiments since the beginning of the space era has been satisfyingly direct and simple. These investigators, often working with <u>co-investigators</u>, (1) plan and develop the experiments, including their instruments, (2) assist in the instrument integration, launch, and in-orbit operation, (3) receive and analyze the data, and (4) publish the results. The very earliest information systems used by these investigators were simple by today's standards. The spacecraft had either no or a very limited ability to respond to ground commands, so the decision-making process did not require the investigators' day-to-day interaction in spacecraft operation. The data were recorded on analog tape at the data acquisition stations, and the original tape recordings or copies were mailed to the investigators for data reduction and analysis within their laboratories. Even in the earliest days of the space program the character of the experiments changed rapidly, and this class of information system was quickly outgrown. The instruments quickly became more demanding in terms of required operational attention, returned data volume, and ground processing complexity. In spite of the fact that it is not as simple and direct as it once was, this principal investigator mode remains the preferred one for many types of experiment.

In the second investigator model, a number of experimenters collaborate on an essentially coequal basis, with one of their number serving as team leader. In its simplest form, principal investigators associated with a particular spacecraft work closely together in this manner during periods of intense coordinated activity such as planetary encounters. Their purpose is to optimize spacecraft operation and data acquisition in the face of the inevitable conflicts, and to stimulate initial scientific productivity. This investigator team model took a more advanced form in the Atmospheric Explorer project, where agreement on the sharing of a common data set made the data from the entire array of instruments equally accessible to all. In addition, the central data handling system accommodated some of the analysis processing, so that the researchers were able to share intermediate results and processing algorithms. The primary motivation was to facilitate and stimulate the collaborative examination of broader problems within the aeronomy discipline than might have been possible by independent efforts. Investigator teams place new demands on their information systems in terms of operational responsiveness, rapid data delivery, and a need to process data from multiple sources.

Several programs now include specific provisions for guest investigators (sometimes referred to as guest observers). These investigators do not necessarily participate in the initial mission planning or instrument design phases, but have the opportunity to propose and conduct investigations with NASA support through the use of active spacecraft and the data acquired from them. This model is similar to the one in which researchers compete for observing time on major ground-based telescopes. In space, for example, the Hubble Space Telescope plans a major guest investigator component. To support guest investigator programs, the information systems must have many of the attributes of those outlined above for investigator teams, e.g., responsiveness, flexibility, and regular and timely data delivery.

As the broad disciplinary and multidisciplinary studies mentioned earlier become more common, we can expect the evolution of the researcher consortium model. Here, a broad problem area is selected which requires the participation of investigators having a wide range of training and experience. Observations are required from a variety of sources, such as multiple instruments, multiple spacecraft, aircraft, and ground sites. The unique relationship between a specific spacecraft-borne instrument and an associated line of investigation disappears. For these studies the information systems must have an even greater ability to deal with diverse data sources, a variety of processing algorithms and models, and widely dispersed collaborative investigator sites.

Finally, many investigators make use of the data after the initial acquisition and analyses phases. These retrospective investigators usually play no part in instrument and data system development and operation for the missions from which they use the data. They may include, in addition to researchers in the usual space and earth sciences disciplines, such diverse individuals as, for example, city planners, agricultural experts, lawyers, authors, and school teachers who customarily use any easily available information in their work. These uses of the data are relatively intangible; it is often difficult to design the systems to meet these initially unspecified needs. These retrospective investigators are, nevertheless, a valid and important segment of the entire information-using community. They are best served by well designed, documented, and relatively stable archives. It is this segment of the user community which has been served least well in the past.

As the possibility of the expanded use of space-acquired data to support both federal agency operational responsibilities and the large communities of other users was recognized, some spacecraft came to be operated as general purpose facilities. Polar orbiting and geostationary environmental satellites are operated in this manner by the National Oceanic and Atmospheric Administration. These spacecraft provide data in near-real time to the National Weather Service and National Ocean Service forecasters and other environmental data users. Various forms of the operational data are captured and made available for longer-term research. This wide variety of uses places an extra heavy burden on the operational satellite information systems. Finished operational data products must be delivered on a regular "production line" basis, with time delays which range from minutes (for environmental warning) to a few days. Concurrently, the data must be prepared and archived for the longer-term research.

OBSERVATION: Each of the investigator models (principal investigator, investigator team, guest investigator, researcher consortium, and retrospective user) contains features which serve particular needs. Each will persist into the future, including the space station era. The information systems should continue to evolve to serve this variety of researcher environments.

2.4. System Concepts

Information systems for space and earth sciences research have been discussed at length in the two National Academy of Sciences reports by the Committee on Data Management and Computation [NRC, 1982], [NRC, 1986]. The present state of data accessibility is discussed in an enlightened manner in the report by the Academy's Joint Data Panel [NRC, 1984]. Some information system concepts for the space station era are outlined in Section 5 of the report prepared by the Task Force on Scientific Uses of Space Station (the "Banks Committee") [NASA, 1985]. Information system needs for Earth Observations in the space station era have been examined by the Earth Observation System (Eos) Science and Mission Requirements Working Group, the subsequent Science Steering Committee, and its Data Panel [NASA, 1984b], [NASA, 1986]. In the light of the wealth of available literature, this paper limits itself to highlighting a few of the most important system concepts. For an elaboration of these concepts, and for further justification of their need, the reader is referred to the papers cited.

2.4.1. Basic Principles

One of the important contributions of the CODMAC first report is the identification of some general principles which should guide information system development and operation [NRC, 1982, pp. 135-139] . The summary statements from that report are as follows:

"I. SCIENTIFIC INVOLVEMENT: There should be active involvement of scientists from inception to completion of space missions in order to assure production of, and access to, high-quality data sets. Scientists should be involved in planning, acquisition, processing, and archiving of data.

"II. SCIENTIFIC OVERSIGHT: Oversight of scientific data-management activities should be implemented through a peer-review process that involves the user community.

"III. DATA AVAILABILITY: Data should be made available to the scientific user community in a manner suited to scientific research needs and have the following characteristics:

"1. The data formats should strike a proper balance between flexibility and the economies of nonchanging record structure. They should be designed for ease of use by the scientist. The ability to compare diverse data sets in compatible forms may be vital to a successful research effort.

"2. Appropriate ancillary data should be supplied, as needed, with the primary data.

"3. Data should be processed and distributed to users in a timely fashion as required by the user community.

"4. Contractual obligations by users to return data to the archive in a modified form should be enforced.

"5. Proper documentation should accompany all data sets that have been validated and are ready for distribution or archival storage.

"IV. FACILITIES: A proper balance between cost and scientific productivity should govern the data-processing and -storage capabilities provided to the scientist.

"V. SOFTWARE: Special emphasis should be devoted to the acquisition or production of structured, transportable, and adequately documented software.

"VI. SCIENCE DATA STORAGE: Scientific data should be suitably annotated and stored in a permanent and retrievable form. Data should be purged only when deemed no longer needed by responsible scientific overseers.

"VII. DATA-SYSTEM FUNDING: Adequate financial resources should be set aside early in each space project to complete data-base management and computation activities; these resources should be clearly protected from loss due to overruns in costs in other parts of a given project."

2.4.2. Space Science Data Management Units

The Space Science Data Management Unit (SSDMU) concept was introduced in the CODMAC first report [NRC, 1982]. These SSDMUs are defined simply as groups (with their equipment, software, databases and other support activities) which are involved in space and earth science data-related activities. They range in scope from individual researchers in university, government, or other laboratories; to aggregations of investigators in those laboratories; mission centers where investigators and investigator teams work with their data and participate in spacecraft operations; data processing facilities which prepare data for use; larger centers such as the Space Telescope Institute where principal investigators and guest observers work; and data centers.

Each SSDMU is designed and run to meet the specific needs of one or a cluster of participants, and handles one or more data sets in the process. A number of these related SSDMUs, in combination with the associated space-based instruments, operating facilities, and communications facilities, constitute a Space Research Information System (SRIS).

A SRIS includes all of the data-related capabilities required for a particular mission or area of study. Thus, the SRIS for early Explorer spacecraft, operating within the classical principal investigator model, included as SSDMUs the spacecraft control center, the central data processing facilities at the Goddard Space Flight Center, the data processing and analysis facilities of the principal investigators, and the National Space Science Data Center at Goddard. In this relatively simple example the functions of the SSDMUs were very nearly autonomous. That is, although there were linkages between them, including the handover of data from one to another, there was relatively little feedback, and each SSDMU operated with considerable independence on a day-to-day basis.

The systems now range upward in complexity to include those having many SSDMUs which are highly dispersed geographically, with a large amount of frequent two-way interaction between them.

2.4.3. Repositories, Active Databases, and Archives

Each Space Science Data Management Unit (SSDMU) deals with data sets, i. e., collections of data in forms which can be addressed and manipulated. Many of these data sets represent assembled, organized, and stored data which may be useful to other than the original investigators.

The second CODMAC report [NRC, 1986] classifies these databases as repositories, active databases, and archives. <u>Repositories</u> are buffers for new data; they may take the form of mission repositories or instrument repositories. The data are distributed from the repositories to those associated with a specific mission or instrument for their analysis, both for mission operation and for the derivation of first science results. Many of these repositories are active for only as long as the mission is operational. Some of them are warehoused after they have been provided to the investigators; even so, they become increasingly inaccessible with the passage of time because the equipment, software, and knowledge necessary for their use disperses. Such warehoused repositories are usually inadequately structured and documented, and do not have the organizational commitment required, to ensure their usefulness for serious research over an indefinitely extended period of time. Unless specific arrangements are made to turn these repositories over to properly supported and committed permanent archives, they cease to exist for most practical purposes soon after the end of the mission operational phase.

Active databases are defined as databases used in on-going research. They are generally under the direct control of the investigators and are usually housed in their facilities. They often include ancillary data and data from multiple instruments and missions. They frequently outlive the originating missions' operational phases and the direct association with individual researchers, remaining active for as long as a given line of research by the locally associated group is active. Their form is more highly variable than the other databases, as the data are repetitively manipulated, and as correlative and ancillary data are added. They are often relatively poorly documented because of their dynamic nature, and therefore not easily used by those outside the immediately associated investigator groups. They may have an intrinsic value to others, but capitalization on that broader potential demands either greater than normal care by the originators with respect to data structuring and documentation, or highly collaborative efforts between the parent and outside investigators.

The longer-lived data collections known as archives are housed in, maintained, and made accessible by formally established data centers. To be useful these archives should have several attributes. First, the data must have an intrinsic value beyond their initial use. Second, they must be documented and organized in such a way that they can be understood and used by other than their originators. Third, they must be accompanied by appropriate related ancillary data. Fourth, it must be possible for future users to find and retrieve them by the use of directories, catalogs, query, retrieval, distribution, and other supporting functions. And fifth, there must be an adequate organizational commitment and funding base to protect and maintain both the data and the servicing functions for an indefinite, but extended, period of time.

Considerable confusion has existed about these three distinctly different types of database. A common mistake has been to aggregate them when talking about the "archive problem". The CODMAC expended considerable effort to distinguish between them, and to describe their separate characteristics as a basis for objective system planning [NRC, 1986].

2.4.4. Distributed Information Networks

Another conclusion of the two CODMAC reports [NRC, 1982],[NRC, 1986] is that the Space Science Data Management Units, with their repositories, active databases, and archives, should be viewed as elements of a coordinated, geographically distributed information system. They recommend that NASA aggressively pursue an evolutionary approach to a communications network which would interconnect these SSDMUs. The central theme of the National Research Council's Joint Data Panel report [NRC, 1984] is the need for establishing a central data catalog and data access network to serve the needs of the solar-terrestrial research community. The Earth Observations System Report of the Data Panel [NASA, 1986, p. v] states that "The core of the Eos data and information system should be an electronic information network, allowing access to the full suite of Eos system capabilities. This network should be flexible, providing access to mission operations, archives, selected active data bases, and to, for example, large mainframe computers for certain, very intensive, computational activities (e. g., modeling) needed to support Eos data analyses."

2.4.5. Telescience

The concept of a cohesive information network has been carried further by those seeing the merits of a new level of highly responsive interaction between the researchers and their instruments, databases, data processing facilities, and collaborators. It should be possible to link the flight instruments in near-real time with the investigators in their home laboratories, or at centers for disciplinary or multidisciplinary collaborative research. This tight coupling between the users and their instruments, with little or no outside intervention or delay, would permit the kind of highly interactive observation in space which is common today at large ground-based research facilities. In this environment, experimenters would be able to operate their instruments in space as though they were nearby. The term <u>teleoperation</u> is sometimes used to refer to this concept.

This concept can be carried even further to encompass the data handling required for scientific analyses. Investigators can accomplish many of the data-related tasks involved in their analysis process in a highly interactive, dynamic manner. Thus, in addition to controlling the detailed operation of the instruments in space, the investigators can also locate and obtain access to related data, control the data processing, freely exchange software and data sets with collaborators, and in other ways optimize the data distribution, processing and analysis methodology. This process is sometimes called <u>teleanalysis</u>. The broader term <u>telescience</u> refers to the combination of teleoperation and teleanalysis.

Some aspects of this highly interactive mode of operation have already been employed by several flight projects. But under the stimulation of the current critical examination of ideas for the Space Station, the concept is being pushed more vigorously than might otherwise be the case. It is discussed at some length in the Space Station Summer Study Report [NASA, 1985].

OBSERVATION: There is increasing agreement that space research data management planning should be treated as a system problem, involving the concept of distributed Space Science Data Management Units with a hierarchy of databases and integrating networks.

2.5. Data and Software Standards

Many of the data management studies have included discussions of data standards, i. e., formally defined formats or structures for databases, communications links, system interfaces, data directories, data catalogs, and their documentation. Some of these discussions have, in addition, encompassed software design standards, and the adoption of standard computer languages. The most important arguments for these standards are that they would reduce the investigators' workloads and assist in controlling costs.

It is concluded that such standards, and perhaps more importantly, less rigid guidelines, may be important where (1) the expenditure of substantial resources is involved, (2) there is intrinsic value beyond the initial uses, and (3) transferability for other uses is practical. These three attributes have been recognized for certain of the space research <u>data</u>. The extent to which <u>software</u> has value beyond its initial use, and is transferable, has been frequently argued. There appears to be growing agreement that some algorithms (with their resulting software) used for frequently performed functions does possess those attributes.

The international Consultative Committee for Space Data Systems, with strong participation by NASA's Office of Space Tracking and Data Systems, has made substantial progress in developing standards related to the space communications links. But few recommendations for data guidelines and standards related to the other parts of the information systems, particularly those which manipulate the research data, survive beyond their initial recommendation. Although many researchers recognize the potential benefits of such guidelines and standards, they are usually hesitant or unwilling to commit to them for fear that they will prove to be overly burdensome. The CODMAC second report [NRC, 1986], makes a major attempt to place standards in a meaningful context from the scientists' point of view. It makes specific recommendations for NASA to work cooperatively with the research communities to develop standards in the areas of information system interfaces, formats for transportable data, utility software, directories and catalogs, and documentation.

2.5.1. Key Standards Issues

The two classical standards issues deal with their scope, and the manner in which they are imposed. Concerning scope, there are two opposing views, the "few is best" versus the "more is better" approaches. The former, (traditionally held by the investigators), is that the best research progress can be made by adopting only those format and software standards which have already gained widespread acceptance by virtue of their clearly demonstrated superiority. The other point of view, (held predominantly by the project managers and data system designers), is that data system standardization should simplify the overall system design process, make the systems more reliable, save money, and facilitate the more complete use of the data. This latter reasoning is often extended to include the creation of new standards where suitable ones do not exist.

The second major issue has to do with the manner in which data standards are imposed. Again there are two factions, the "trickle-up" and the "impose from the top" camps. The former argues that the best (the only really workable) standards are developed at the working level, spread laterally, and eventually become accepted higher in the organization as their obvious merits become recognized. The second argues that standards will never be used to the extent that they should be if we wait for the trickle-up process; the only effective approach is to impose them from above.

Some degree of standardization has become more desirable as the data have been increasingly shared by multiple participants. It is being driven forcefully by the space station planing.

2.5.2. Areas for Meaningful Guidelines and Standards

Some interfaces are already standardized at the project level as a matter of practicality. Others will become NASA-standard as a matter of course because they involve the general purpose communication circuits. There are additional areas where data guidelines and standards are needed to facilitate and stimulate the wider sharing of data and processing techniques. NASA-wide guidelines and standards are urgently needed for data center directories and catalogs, and for the means for accessing them. It is also time to begin serious standards work for the repositories and the archives, including their documentation. Other major system interfaces should be critically examined; highest priority should be given to the interfaces between Space Science Data Management Units.

The standards issue is especially contentious when it involves software written and used by the investigators. Here, there are four separate arenas where software guidelines and standards, including the choice of standard computer languages, might be considered, (1) software for instrument checkout and calibration, (2) software for in-orbit operation, (3) software associated with investigators' access to the various databases, and (4) software for data analysis. Clearly there must be a degree of standardization in areas 1 and 2, and such standards are often imposed by individual projects. The possibility of adopting NASA-wide guidelines and standards in these areas should be carefully examined. The controlling objective should be to reduce investigator effort and costs by making it possible for them to use the same checkout and operations software throughout the entire mission, from instrument development to in-orbit operation.

For database access (area 3), it is envisioned that several concurrent standards might be supported to accommodate the main bases of researcher experience with multiple computer types, operating systems, and program languages. In addition, a library of utility packages for frequently required functions should be established for the benefit of the investigators.

Attempts should not be made to standardize the programming languages and investigator-developed software for analysis (area 4), at least at this time. This software is so much an integral part of the investigators' research process that its widespread portability is probably an unrealistic goal. It is appropriate, again, to establish a library of utility software routines frequently used by investigators for analysis.

OBSERVATION: Data format, system interface, and documentation guidelines and standards are urgently needed for the directories, catalogs, and their inquiry networks to assist users in acquiring information about data. The development of guidelines and standards for the repositories and archives is also desirable. NASA communication circuit format standards are in existence or being developed under the leadership of the Office of Space Tracking and Data Systems; comparable guidelines and standards are needed for the interfaces between many of the other Space Science Data Management Units. The adoption of guidelines and standards for some of the software associated with instrument checkout and operation, data distribution, and databases should be critically examined. Selected documentation guidelines and standards are also needed. The imposition of standards within the investigators' analysis facilities should not be attempted at present.

RECOMMENDATION:

It is recommended that the Office of Space Science and Applications, working closely with the NASA field centers and external research communities, prepare a plan for the evolutionary development and promulgation of carefully selected data format, system interface, software, and documentation guidelines and standards. This plan should encompass selected Space Science Data Management Unit interfaces, repositories, archives, data directories, data catalogs, and their documentation. It should include the establishment of libraries for exchanging frequently used investigators' utility software. Responsibilities should be clearly assigned for developing and carrying out this plan.

2.6. New Technology

The CODMAC second report [NRC, 1986] contains a good overview of many of the developing technologies which are related to future growth in space research data management. The reader is referred to that document for detailed information. Its Executive Summary includes the following:

> "A wide range of hardware, software, and systems technologies is becoming available and must be considered by NASA and the space science community in moving toward an information system involving coordinated sets of distributed SSDMU's. For SSDMU's involving small research groups, simple workstations, consisting of microprocessor-based terminals, with modest storage devices and other appropriate peripherals may be sufficient. Other groups will require minicomputers and a number of special purpose peripherals. On the other hand, SSDMU's with a charter for pipeline processing of significant quantities of data, or for long term maintenance of data, will require more sophisticated and capable data handling, processing, and storage systems. Solutions to computation and data management problems must thus be flexible and dependent on the type of SSDMU being considered.
>
> "<u>We recommend NASA's posture in terms of new computation and data management technologies be largely to maintain an awareness of advances, and to assist the space science community and associated SSDMU's in utilizing the technologies in meaningful ways.</u> Awareness is the key because most technology advances in computation and data management will probably arise through industry-supported activities. There is an important area where unique requirements of the space science community suggest technology development efforts are needed. That area deals with development of portable software packages that are designed for wide use in the space science community. Portability means developing software in higher level languages, in reasonably machine-independent form, and with use of acceptable standards. Widespread use of such packages should alleviate some of the current problems in transfer of data between SSDMU's and should facilitate distributed archiving and processing of data. An expert systems approach to analyses of imaging spectrometer data, and coupling of advanced database management software with spatially and temporally tagged vector and array data, are but two examples of needed software development efforts that are not being vigorously pursued by industry at this time."

OBSERVATION: The lack of appropriate basic technologies does not appear to be a major factor limiting present space research data use. The diligent application of available technologies, coupled with suitable planning, coordination, and management, could satisfy most of the needs for the

presently approved missions. Carefully selected technology advancement is needed, however, for some of the longer-range future missions, including some of those envisioned for the space station. Areas needing attention include (1) on-board data processing, (2) distributed interactive networks for science operations and analysis, (3) very large mass data storage, (4) database structures, including the use of database machines, (5) transportable software, and (6) the application of artificial intelligence approaches. NASA needs to develop and maintain a keen awareness of technology developments in these areas. In selected instances, NASA should fund technology development which it needs, and which would not otherwise be forthcoming.

3. DATA CENTERS

From the earliest days of the space era there has been a strong belief in the importance of archiving data to make them available for future researchers. This continued a national and international data center tradition which was greatly strengthened during the late 1950's as a part of the United States contribution to the International Geophysical Year. As a result, the National Space Science Data Center (NSSDC) was established at the Goddard Space Flight Center early in NASA's life.

3.1. The National Space Science Data Center

The original role of the NSSDC was to collect the science data from the nation's rocket and satellite programs and make them available for later analysis by other than the originating principal investigators. Since then it has been recognized that the NSSDC could fill additional important roles in further stimulating the increased use of the science data.

The National Space Science Data Center has been a long-standing object of debate and concern. Throughout much of its history there has been criticism that it was not fulfilling its envisioned role. This role was expressed in NASA Policy Directive 8030.3 dated January 7, 1967 and further defined in NASA Management Instruction 8030.3A, dated May 2, 1978 (see Appendix C). Among other things, these documents established the specific responsibilities of the Director of the NSSDC and of the NASA program, project, and management officials for planning and executing flight missions so that the mission data would be appropriately handled and ultimately archived. The latter Management Instruction called for the preparation of a Project Data Management Plan for each project to spell out the arrangements and responsibilities for, among other things, archiving the space research data. Further, it outlined the responsibilities of the NSSDC for supporting investigations by making its scientific data and facilities available. Appendix B to that management instruction is a detailed description of the intended functions of the NSSDC.

These functions have not always been performed as envisioned. In some cases the data have not been acquired by the Data Center; in others, data have been collected, but not in forms which made their use by third party investigators easy or, in some cases, possible. Some attribute this to the low level of support for the NSSDC by management, including inadequate budgets. Others cite failure of the project and program managers to enforce the requirements of NASA Management Instruction 8030.3A. Still others attribute it to failure of the NSSDC to take an energetic leadership role. And the principal investigators have been cited for their failure to satisfy their contracted responsibilities to place their data in the archives. All of these factors undoubtedly contributed to the situation as it existed a few years ago.

In spite of this background, there is a growing belief that carefully preparing and archiving correctly chosen data sets and providing easy access to them is important. To some extent this increasing emphasis may be a consequence of the space research maturation process described in Section 2. This need is discussed at length in the two CODMAC reports mentioned earlier [NRC, 1982], [NRC, 1986], in the Joint Data Panel report [NRC, 1984], and in the Earth Observation System Data Panel report [NASA, 1986].

A recent organizational restructuring at the Goddard Space Flight Center was undertaken to, among other things, strengthen the National Space Science Data Center, and good progress is being made in this direction. There is also a growing recognition within NASA that data sets of comparable worth reside at other field centers, and are in danger of loss from lack of adequate provisions for their care. For example, some of the planetary data sets presently stored at the Jet Propulsion Laboratory represent a unique national resource and they, too, should be properly organized and safeguarded for future use.

3.2. Suggested Changes

A number of steps are suggested to make the data center activities more viable.

3.2.1. A Distributed Data Center Concept

An active, distributed network concept should be adopted to make the data centers a more dynamic part of the research process. This should include the National Space Science Data Center, but should also encompass holdings at other locations. These other holdings, such as the planetary data sets at the Jet Propulsion Laboratory, should be designated as official NASA archives just as are those physically located at the NSSDC. The infrastructure necessary to link these holdings and assure their long-term quality and accessibility should be established. It is immaterial from the user's point of view whether the NASA space research archives are organizationally aggregated or exist as independent data centers, as long as they are associated by a central directory, linked catalogs, guidelines and standards, similar or compatible access and funding arrangements, and unified and cohesive system planning and overview.

3.2.2. Directory and Catalogs

High priority should be given at the NSSDC to the building, maintenance, and operation of a single, centrally administered directory and inquiry system. This directory and inquiry system should be designed to inform potential users of the existence and locations of data of interest. It should encompass the data in all appropriately designated NASA data centers, as well as provide a referral service to major related holdings at other government agencies and universities. It should be possible for a user anywhere in the world with access to a telephone line and simple terminal to call one well-publicized number to gain access to this central directory and inquiry system.

Each data center should maintain high-level and detailed catalogs covering its data holdings. The potential user, having determined the availability and location of potentially useful data from the central directory and inquiry system, should be able to call upon the catalogs for a further search by the use of commercially available terminals, standard line protocols, and uniform inquiry commands. These catalogs should contain enough additional information to permit the users to decide whether they want to acquire the data, and how to do so. A further desirable design goal is for the user to be able to determine the cost of the data and to order them before ending the inquiry session. Consideration should be given to providing direct access to at least the higher level catalogs through the central directory and inquiry telephone connection.

3.3.3. Acquisition and Purging Decisions

NASA should establish an explicit, formal process by which decisions are made as to which data sets should be acquired by the data centers, and which data sets should be purged. The research scientists associated with the data sets should be active partners in this process.

3.3.4. Funding and User Charges

It is essential that viable funding arrangements be worked out for all of NASA's data centers. As a general guide, and in agreement with the deliberations on this subject by the National Research Council's Geophysics Study Committee, the following three-tiered financial structure is suggested:

- The project offices which fund data generation should carry the full responsibility for placing their data into the data centers. The projects should be planned from the beginning with the long-term archiving needs in mind (as is required in NMI 8030.3A), and their budgets should provide for the data validation, formatting, cataloging, documentation, and other data preparation which may be necessary to make the data generally usable in the data center environment.

- Funds should be appropriated directly to the data centers for (1) obtaining and maintaining capital equipment associated with the archives, (2) maintaining the data sets (including regeneration as required), (3) building and maintaining the directory and inquiry system, (4) maintaining the catalogs for holdings which are under their care, (5) supporting the inquiry and data exchange networks, and (6) applying new technologies as required.

- User charges should, in general, cover the cost of extracting data from the archives, preparing special products, and sending the products to the users. Funds should be appropriated to the data centers to cover those services which are exempted from user charges (see below).

It should not be necessary for investigators supported by NASA contracts to pay the data centers directly for the purchase of data needed for their investigations. As a general rule, the investigators should be provided with all the data needed for their work as a part of their contract. Their needs for such data should be included in their proposals, discussed during contract negotiation, and written into their contracts (as is currently required by NASA Management Instruction 8030.3A). The headquarters program offices should provide block transfers of funds to the data centers to cover the total costs of data thus required by the aggregation of investigator contracts.

There are data users for whom charges might be waived. They might include, for example, investigators who are collaborating with, but not funded by, NASA. Funds to cover these users should be either provided by the cognizant headquarters program office, or included as a part of the data centers' direct funding bases.

Consideration should be given to establishing multiple fee schedules. Commercial users should be charged the highest price. Non-profit organizations, such as universities, municipalities, etc., might be charged a lesser fee. And completely independent, unsupported users (the high school student working on a school project, for example) might be charged a minimal or no fee. The primary motivation in establishing the fee schedules should be to maximize the data use at minimum cost to the government.

Action is needed to extract greater agency and scientific benefit from the user charges. The present NSSDC interpretation of Bureau of the Budget Circular A-25 permits them to waive the charges for most users, and a majority of the charges are, in fact, waived. One reason for this is that the data centers presently derive no benefit from the collected fees, since they go directly to the United States Treasury. Several specific steps are suggested:

- Arrangements should be made for the data centers to retain the fees which they collect to help cover the cost of data preparation and distribution. Either the necessary authority should be obtained for NASA to use the fees directly (the National Oceanic and Atmospheric Administration has this authority for its data centers), or these specific functions should be contracted, with the contractor collecting and using the fees to help cover its cost of operation.

- The data dissemination end of the data center activities should be emphasized and conducted more like a commercial enterprise, with a marketing function and a separate accounting system.

3.3.5. Oversight and Evaluation

The entire data center activity (the long-term archives and their support infrastructures) should be under the headquarters purview of the Director of the Information System Office, acting for the Associate Administrator for Space Science and Applications.

The data center activities should be reviewed along with the other space research data management activities by the peer review process recommended in Section 6.5.4.

RECOMMENDATIONS:

1. It is recommended that the Office of Space Science and Applications, with participation by the field centers and internal and external researchers, develop a cohesive plan for the data center activities, including not only the National Space Science Data Center, but the other data holdings for which long-term retention and accessibility is desired. These entities should be operated as a dynamic, integrated network in a manner which will cause them to play a more active role in the research process. One of the most urgent first steps should be to work out a cohesive, viable funding structure for all data center activities.

2. The NSSDC, in addition to its other functions, should be responsible for (1) establishing and operating a central data directory and inquiry system, (2) leading the establishment of catalog, archive, and documentation guidelines and standards for use by all of the NASA data centers, and (3) arranging for the exchange of catalogs and data with other agencies and countries.

4. POLICY GUIDANCE

It has been argued that NASA might more completely achieve its space research data management objectives if there was a clearer statement of agency policy on the subject. A careful examination of NASA's written space research data management policy reveals that it is in surprisingly good shape. Appendix B lists the most pertinent policy documents. The key document is NASA Management Instruction (NMI) 8030.3A, Policy Concerning Data Obtained from Space Science Flight Investigations, dated May 2, 1978; a copy is included as Appendix C for convenient reference. This document contains all of the essential elements for meaningful policy guidance, specifically:

- There is an understandable statement of the document's purpose.

- There is a concise, clear statement of the policy, backed by explanatory material to assist in its understanding.

- There is an assignment of responsibilities for implementing it.

4.1. Suggested Changes

There are several ways in which this document might be updated and otherwise improved:

- The NMI was written when most investigations were performed by principal investigators. As was discussed in Section 2 of this report, other modes of investigator participation are growing in importance. Researcher consortia and retrospective users should be added to the user models described in the NMI (principal investigators, team leaders, and guest investigators). The role of scientist-astronauts in manned flights should be described. Data accessibility, data rights, and investigator responsibilities should be covered for guest investigators and the participants in researcher consortia.

- The NMI was written when most data were being distributed via magnetic tapes through the mails. Although mailing data on tape (and eventually on optical media) will continue, electronic distribution by use of commercial circuits and networks is becoming increasingly prevalent. Policy dealing with data accessibility, control, and security in the network environment should be added. Note should be taken especially of the changes in research methodology anticipated for the space station era.

- The NMI treats the National Space Science Data Center primarily as a final resting place for data. This should be changed to include the other functions described in Section 3. As stated in that section, the NSSDC role should be that of an active facilitator of space research.

- The instruction stresses the data acquisition end of the NSSDC's mission, virtually ignoring the data distribution function. Data inquiry, searching, browsing, distribution, and, in fact, active merchandising, should be given appropriate attention.

- The present NMI addresses both policy and implementation details. This may be confusing. Consideration should be given to developing a concise statement of policy as a NASA Policy Directive (NPD), and covering the implementation details and assignment of responsibilities in an update of this NMI.

4.2. Enforcement

With the changes suggested above this documentation could serve as a exemplary model of high quality data management policy guidance! But that would not necessarily make it any more effective than it has been in the past. It has lacked impact because many of its provisions have been ignored. Steps need to be taken to enforce its provisions.

RECOMMENDATIONS:

1. It is recommended that the policy elements in NASA Management Instruction 8030.3a, expanded as necessary, be incorporated in a short NASA Policy Directive. The remaining portion of the NMI, including the background, explanatory materials, implementation instructions, and delegation of responsibilities, should be updated to reflect the changes which have occurred, and which are envisioned for the future.

2. It is recommended that the Office of Space Science and Applications take specific steps to enforce the provisions of the resulting policy and implementation guidance.

5. ORGANIZATIONAL STRUCTURE

The National Aeronautics and Space Administration employs a matrix organizational structure. In the formal line structure, programmatic Associate Administrators are responsible for the activities of the Office of Space Science and Applications (OSSA), Office of Aeronautics and Space Technology (OATS), Office of Space Flight (OSF), Office of Space Station (OSS), and Office of Space Tracking and Data Systems (OSTDS). Field centers report to the Associate Administrators for OSSA (Goddard Space Flight Center and Jet Propulsion Laboratory), OAST (Ames Research Center, Langley Research Center and Lewis Research Center), and OSF (Johnson Space Center, Kennedy Space Center, and Marshall Space Flight Center).

The space research programs are managed by program directors in the Office of Space Science and Applications. Current program areas are Life Sciences, Communications, Earth Science and Applications, Solar System Exploration, Microgravity Sciences and Applications, and Astrophysics. These directors are responsible for overall program content in their disciplines, and for the specific flight missions in their areas, including data management. They work out arrangements for support across the various line offices as needed for the accomplishment of their missions.

Individual flight projects are managed at the field centers by project managers and their staffs. They are the ones most directly responsible for building the spacecraft and arranging for the facilities needed to support their mission. They arrange for support from other organizational elements within their field centers and, working through their program directors in headquarters, from other headquarters offices and field centers.

This management arrangement has served NASA well. It permits the establishment of a strong focus on the space flight missions without requiring a major reorganization each time a new mission is undertaken.

One of the natural consequences of this structure is that the direct management responsibilities for different elements of the space research data systems are divided within the NASA organization. The Office of Space Science and Applications, with its supporting field centers, has the overall mission responsibility, and oversees the research spacecraft design, the development of some of the ground facilities (including some of the control centers), the operation of the spacecraft, the data analysis, and the data archiving. The Office of Space Tracking and Data Systems, with its supporting field centers, is responsible for data acquisition, tracking, trajectory and orbit determination, many of the communications circuits, and some of the control centers. The Offices of Space Flight and Space Station, with their field center support, manage the large manned spacecraft with their mission control subsystems. The Office of Aeronautics and Space Technology supports new technology development of interest to all offices.

There is nothing inherently wrong with this distribution of data system management responsibilities. But it does require careful attention to make sure that all of the data system requirements are identified, that responsibilities for all activities are assigned, that the component facilities meet the requirements, and that they work well together. Two specific points deserve further discussion.

First, the strong primary emphasis of the project managers at the field centers is, appropriately, on flying successful missions to meet carefully specified objectives within cost. Their approach to data system design tends to be directly focused on that objective. There are very strong pressures on them to design the minimal data systems which will meet their relatively well-defined science mission objectives. In addition, their natural tendency is to build mission-unique facilities whenever possible, so that they will have the most direct possible control over cost and performance. These factors make it difficult to move in some of the new directions with respect to space research data management which are discussed in Section 2 of this report.

The program directors in the Office of Space Science and Applications are the ones primarily responsible for the decisions which might improve the ability to conduct broader studies within their disciplines, even though this might cost more than simply meeting the immediate objectives of the individual missions. And they are the ones who must take the initiative in making the tradeoffs and arrangements to facilitate collaborative studies which may cut across their traditional discipline boundaries.

The second point concerns broad system planning and the conduct of data system studies. The individual OSSA program offices do not have much of the expertise required to plan data systems or to derive the information needed for their tradeoff decisions. They obtain this support from the Office of Space Tracking and Data Systems in the areas of data acquisition, command transmission, communications, orbit/trajectory determination, and spacecraft operations. But past attempts to obtain comparable support for other than this central portion of the complete data system from other headquarters offices have had marginal success. This is because that support requires a detailed understanding of the investigators' space research processes. None of the other offices has the required depth of understanding of this process, and therefore of the special needs of the investigators.

It was for these reasons that the Associate Administrator for Space Science and Applications established the Information Systems Office several years ago. It was envisioned that this office, with expertise in data systems, but with close ties to the research program offices, would provide the broad information system planning and support needed by the program directors. For reasons discussed in the next section, the Office of Space Science and Applications has been less than fully successful in this role.

It has been suggested from time to time that NASA establish a new headquarters office with its own Associate Administrator to take over all space research data management functions. This approach to integrating the data management functions is not recommended, since it would further diffuse the program managers' abilities to conduct their individual flight missions.

OBSERVATION: The present National Aeronautics and Space Administration organizational structure, in general, serves the agency well. Although the space research data management responsibilities are widely dispersed, there are no fundamental reasons why they cannot be adequately coordinated to provide cohesive, well-integrated data systems to meet present and future needs. It is suggested that only minor changes to the present overall NASA organizational structure are needed for the purpose of improving space research data management.

6. PLANNING AND MANAGEMENT

The process of gaining approval for new flight projects is arduous, culminating in Congressional budget hearings. Only those projects survive for which both the overall program objectives and the specific mission objectives have been clearly stated. The research community assists extensively in this process through its participation in laying out NASA's long range research program, and in developing specific program and mission objectives. Thus, each approved project starts with a clear idea of the specific objectives which it hopes to achieve.

The translation of those objectives into reality requires extensive planning and management. That planning and management as it relates to the information handling activities is the object of this section's discussion.

6.1. Achieving Initial Spacecraft Operation

There is a very strong focus within NASA on the mission activities which culminate in the establishment of initial spacecraft operation. These include instrument and spacecraft planning, designing, building, launching, in-orbit activation, first operation, and first data analysis. They encompasses the development of the hardware, software, operational procedures, and all other elements of the space and ground segments which are required to put the spacecraft into orbit and make them work.

Achievement of first operation is a major milestone in the history of NASA projects. Here it becomes apparent for the first time, after years of hard work, whether the spacecraft are able to provide the quality and quantity of data which was envisioned. It is also the point at which it becomes very difficult or impossible to correct design flaws in the spacecraft subsystems, or to fix many of the failures which may occur. This is the most easily visible demonstration of project success or failure.

NASA has an outstanding record of success in reaching this point. It is primarily for this reason that the agency is so often regarded nationally and internationally as the premier example of a successful high-technology organization, with well-developed skills for planning and managing large and complex technological endeavors. NASA has been able to achieve this success, among other reasons, because it has such a well worked out and diligently applied process for planning and managing the development, construction, and operation of the on-board and ground subsystems needed to control the spacecraft, to track the spacecraft and compute their trajectories, and to acquire their data. Throughout NASA's history only one research spacecraft (a Ranger) has been lost due to deficiencies in performing those functions. In fact, NASA's outstanding performance in these areas has contributed to the salvation of many missions where the spacecraft have malfunctioned in complex and unforeseeable ways, and would have otherwise been lost.

OBSERVATION: NASA has excelled in planning and managing the development of the spacecraft and the directly supporting ground subsystems required for their initial activation.

6.2. Achieving the Mission Research Objectives

The flight program and project staffs are also responsible for operating the spacecraft for their full productive lives, and for analyzing the data to achieve the initially stated mission research objectives. This requires the establishment of an infrastructure and facilities which are just as important (even if not as visible) as the ones required for initial spacecraft activation. In a sense these capabilities are more difficult to plan and manage well because the research process for which they are required is in many cases not as well systematized, and because the participants are so widely geographically, organizationally, and procedurally dispersed.

On the whole, these operational and analysis processes have been quite successful. Throughout the agency's history most of the initially stated research objectives have been met and many discoveries have been made which were unforeseen at the time of mission conception. As in the technology arena, NASA research is generally viewed by the scientific community and the public at large as being highly productive and creative.

There have been criticisms within the research community that those results might have been achieved more quickly and with less frustration and effort, and that substantially more could have been achieved with modest additional resources, if there had been better information system planning and management. The CODMAC first report [NRC, 1982] addresses this issue in considerable detail, and it is examined further in Section 6 below.

OBSERVATION: An amazing amount of new scientific information has been gleaned from the space research missions. A majority of the specific science objectives for the individual missions have been met. It may have been possible to achieve even more of the mission potential through better planning and execution of the data management activities.

6.3. Beyond the Individual Flight Projects

It is in looking beyond the capabilities of the individual flight missions that one sees the greatest opportunity for future gain. If major problems are to be addressed within broad scientific disciplines or across related disciplines, as discussed in Section 2.1, data systems are needed which go beyond the specific needs of individual flight projects. If advantage is to be taken of new approaches to research data management as discussed in Section 2.2, or if some of the new system concepts outlined in Section 2.4 are to be implemented, a more cohesive approach is necessary. If a truly utilitarian space station design is to be achieved, that is, one which can support a wide variety of research endeavors effectively and with ease, then a new kind of information system planning and management methodology is needed. This will require a substantial shift from the traditional concentration on individual research projects.

The staffs of the Office of Space Science and Applications and the associated NASA field centers have taken steps to move in this direction. The concept of pooled data sets and derived products was followed in a number of individual satellite projects, e.g., Atmospheric Explorer, International Sun-Earth Explorer/International Cometary Explorer, Solar Mesosphere Explorer, Solar Maximum Mission, and Pioneer-Venus, among others. This was usually done through the establishment of guest investigator and interdisciplinary scientist programs, where research proposals were solicited from scientists outside the immediate project teams, sometimes for studies requiring data from multiple instruments. Another important step in this direction was the organization of Coordinated Data Analysis Workshops by the National Space Science Data Center.

The Space Plasma Computer Analysis Network linked a geographically dispersed group of Solar-Terrestrial Physics investigators and made a central computer and database available for their work. This was done at relatively low cost from the Marshall Space Flight Center by linking mostly-existing facilities. This network was so successful that it has been expanded to form a broader Space Physics Analysis Network.

The Information Systems Office was formed within the Office of Space Science and Applications to act as a focal point for research data management. One of this Office's initial tasks was to oversee four pilot data systems (the Pilot Climate Data System, Pilot Land Data System, Pilot Ocean Data System, and Pilot Planetary Data System) being developed in recognition of the need for substantial distributed capabilities in specific disciplines. In recent months this office has begun the development of a broad concept for space research data management which has come to be referred to as the Science and Applications Information System.

Part of the rationale for the recent organizational changes at the Goddard Space Flight Center was to provide a better foundation for meeting the needs, not only for research data processing and archiving services in general, but for the multiinstrument, multidisciplinary, large spatial scale, and long duration studies of the future. The Jet Propulsion Laboratory and the Marshall Space Flight Center have also been re-examining their approaches to data system planning and management in recognition of these needs.

The Space Station project has recognized the importance of user instrument operation and science information handling, and has established mechanisms to give the user communities a voice in their planning.

These steps are cohesive in the sense that they are motivated by a common desire to move forward in this area. They are still, however, essentially independent initiatives. A number of flight projects still lack adequate provisions for making their data useful to researchers other than their own directly supported investigators. The pilot data system projects have only relatively recently begun addressing their relationships with each other, with other related projects, and their futures beyond the initial several-year funding. And the space station basic operational philosophy for conducting space research is only slowly being worked out. It appears that NASA information system planning still lacks an overall vision as to the broad sets of requirements the scientists will have for operating their instruments, and for handling, processing, and storing the large quantities of data expected from future missions in the research environment envisioned in Section 2.

OBSERVATION: NASA has taken many specific steps to make improved use of space research data and to facilitate the more complex missions and the broader classes of study presently envisioned. But an overall, cohesive, long-term vision of this data management task, including a realistic understanding at the highest organizational levels of the resources and management attention required, has been lacking.

6.4. Researcher Involvement

A key point made in the CODMAC report [NRC, 1982] is that there must be active involvement by the research scientists in planning and watching over the data management functions if they are to be fully effective. A frequent approach to building data systems has been to assign the planning and development to engineering teams. In some cases the investigators were not invited into the process; in others they chose not to participate. This has led to difficulties, since there are major differences between the motivations and approaches to problem solving of the engineers and the researchers. The engineers may choose approaches which are overly complex, do not satisfy maturing science needs, or lack flexibility to accommodate evolving technologies. The researchers, being preoccupied with the analytical research process and mindful of the pressures for publishing results, wish simply to have the data quickly, with assurances that it has been processed accurately and in an understandable way. As a result of the lack of teamwork between the two groups, the researchers have frequently complained that the system designers did not provide what they needed, while the engineers argued that the researchers would not help them understand their needs. In instances where there has been an active partnership between the researchers and engineers, the systems have worked well.

OBSERVATION: To be effective, space research data management systems must be planned, built, and operated through a active partnership between the technical professionals and the researchers. NASA has an obligation to provide opportunities for meaningful researcher participation in planning and building the systems. In turn, the researchers have an obligation to devote the time and energy required to actively assist.

6.5. Office of Space Science and Applications Roles

The ability to conduct the types of advanced investigation described in Section 2.1 will depend upon careful attention to all parts of the data handling systems and their interfaces so that data from multiple instruments and missions, used with processing algorithms devised by researchers at different institutions, and acquired over a long period of time, can be used effectively. To achieve substantial advances in space research data management it is necessary that a number of specific planning, oversight, and evaluation functions be performed. These functions are a proper responsibility of the Office of Space Science and Applications; no other office understands the research process or has the regular contact with the researchers required to successfully perform them.

6.5.1. Planning

The lack of overall information system planning and management is a recurrent theme in many of the reports, and in numerous of the discussions held during the course of this present study. As was discussed in Section 5, the planning, budgeting, and managing responsibilities for different parts of the information systems are spread throughout the NASA organization. It has been difficult for the project managers to coordinate all elements of the system, and sometimes impossible for them to make the tradeoffs which might have improved their overall research productivity.

Many individuals throughout NASA have recognized this problem and attempted to address it. During the past few years there have been frequent references to "end-to-end" system planning and design as one approach. There remains, however, considerable confusion as to what that means and how to achieve it in the existing organizational structure. In general, NASA does not have a well-established body of experience or management methodology for bringing about integrated and cohesive information system planning and management across the organizational boundaries.

There remains a nearly complete absence of substantial long-range strategic data management planning in NASA. In this context, the term <u>strategic</u> connotes the coordinated employment of all of NASA's applicable resources, through large-scale, long-range planning and development, to achieve broad data management objectives as discussed in Section 6.3. It is distinct from the term tactical, which means the achievement of the more limited mission objectives discussed in Sections 6.1 and 6.2. Strategic planning is essential as the basic foundation for the larger-scale, longer-term studies discussed in Section 2.

It is important that the senior executives in the Office of Space Science and Applications recognize this need, and assign responsibilities and allocate sufficient resources for addressing it. It is not necessary that a large planning staff be assembled within headquarters for this purpose; much of the technical work can be done by participating field centers. Improved high-level leadership is needed in the OSSA, however, to:

- Orchestrate the development of a long-range, strategic plan for space research data management which can serve as a general guide for data-related facility development in the future.

- Lead the development of a philosophy and plan for a cohesive space research data directory, catalog, and related inquiry system, including the attendant guidelines and standards.

- Establish an organized process for identifying new systems approaches for space research data management. The increased use of data networks and telescience are examples of such changes.

- Ensure that the internal and external researcher communities are actively involved in NASA's data system planning.

6.5.2. Headquarters Level Oversight

Responsibilities for carrying out the elements of NASA's programs, appropriately, are delegated to the various headquarters offices and field centers. There is a pressing need, however, for central oversight over certain aspects of space research data management. This oversight function should:

- Oversee the preparation, and follow up on the implementation, of broad data management policy.

- Assemble and defend the budgets for the research data facilities which are outside the scopes of the discipline program directors and individual flight projects. This includes, most importantly, the budget for the National Space Science Data Center.

- Assist the program directors in the preparation and periodic review of the data-related portions of their OSSA Research and Development budgets, including the allocation of funds for the data management functions and changes to those allocations. Advise the Associate Administrator for Space Science and Applications with respect to those budgets and allocations to ensure compatibility with policy and the long-range strategic data management plan.

- Assist the program directors in the periodic review of the planning, design, and operation of the flight project and discipline-oriented data facilities to ensure compatibility with data management policies and the long-range strategic plans.

- Review throughout NASA the requirements, plans, and actual arrangements for services to support the OSSA objectives (for example, data acquisition, tracking, communications, and spacecraft control centers) to ensure compatibility with the individual mission and long-range strategic plans.

- Provide the headquarters level oversight for the large, general purpose computers, databases, and communication networks which are established by the Office of Space Science and Applications to serve multiple programs or projects. Ensure the establishment of explicit arrangements for access to those facilities by participating investigators throughout the internal and external research communities.

- Provide headquarters level oversight for the pilot data system projects up to the point where they are either taken over by the program directors or terminated.

- Working closely with researchers and the standards activities of the other headquarters offices, lead in the selection and promulgation of appropriate guidelines and standards for space research data system interfaces, databases, data directories, catalogs, software, and their documentation which will make it easier for investigators to work more effectively with data from multiple sources.

- Facilitate the exchange of data across agency and national boundaries.

6.5.3. Advanced Development

There also needs to be a central focus for ensuring that the future data management-related technology needs of the Office of Space Science and Applications will be met. This function should:

- Maintain an awareness of, sponsor tests, and otherwise evaluate new and emerging technologies which offer promise of improving NASA's space research data systems.

- Identify information technology areas where development is needed to meet future space research goals. Make them known to the Office of Aeronautics and Space Technology and to others who may be able to contribute to such development.

- Participate in the periodic review of NASA's technology development plans and programs to help ensure that future Office of Space Science and Applications information management needs can be met.

6.5.4. Evaluation

There needs to be a systematic, periodic process for examining NASA's space research data management activities to assist in detecting and correcting deficiencies and advancing future system improvements. The Office of Space Science and Applications should:

- Set up an annual peer review process to assess space research data management performance and assist in identifying desirable changes, including new directions for the future. This task might be assumed by an existing entity such the NASA Advisory Council; alternatively, the National Academy of Sciences/National Research Council could be asked to perform this function.

- Identify a point of contact for receiving and acting upon the inputs from outside-NASA data-related evaluations, such as the National Academy of Sciences study reports.

OBSERVATION: The absence of cohesive long-range strategic planning has been a persistent aspect of NASA's space research data management. There is a pressing need within the Office of Space Science and Applications for improved leadership in this planning, in program oversight, and in performance evaluation.

RECOMMENDATIONS:

1. It is recommended that the Office of Space Science and Applications undertake, as a high priority task, the development of a long-range strategic plan for space research data management. This plan should encompass all phases of data handling and analysis from instrument checkout and mission operation to data archiving, and span the entire system from the sensors through the databases. It should concentrate on general principles

and capabilities, rather than detailed designs. Preparation of the plan should actively involve the headquarters discipline directors, program managers, field center project personnel, information system professionals, and external research communities. The plan should embody the system concepts and the principles for successful scientific data management outlined in Section 2.4. Once approved, this plan should serve as the basis for specific mission and discipline data system planning and budget formulation.

2. It is recommended that the Associate Administrator for Space Science and Applications assign to designated officials specific responsibilities (with accompanying resources, including field center support) in the areas of data management long-range strategic planning, data policy, guidelines and standards, data system budgeting, oversight, advanced development, and performance evaluation, as outlined in the text.

3. It is recommended that an external peer review group be charged with annually (1) reviewing NASA's progress in information system planning and implementation, (2) evaluating system performance, (3) assessing conformance with data policy directives, and (4) suggesting changes and new directions for the future. The review group membership should be roughly balanced between practicing space researchers and information system professionals.

LIST OF ACRONYMS

AE	Atmospheric Explorer
AMPTE	Active Magnetosphere Particle Tracer Explorer
ARC	Ames Research Center
CODMAC	Committee on Data Management and Computation
DE	Dynamics Explorer
EOIS	Earth Observation Information System
Eos	Earth Observation System
HEAO	High-Energy Astrophysical Observatory
IMP	Interplanetary Monitoring Platform
ISEE	International Sun-Earth Explorer
IUE	International Ultraviolet Explorer
NASA	National Aeronautics and Space Administration
NMI	NASA Management Instruction
NRC	National Research Council
NSSDC	National Space Science Data Center
OAST	Office of Aeronautics and Space Technology
OSF	Office of Space Flight
OSS	Office of Space Station
OSSA	Office of Space Science and Applications
OSTDS	Office of Space Tracking and Data Systems
SME	Solar Mesosphere Explorer
SMM	Solar Maximum Mission
SRIS	Space Research Information System
SSDMU	Space Science Data Management Unit

REFERENCES

[NASA, 1984a] The Pilot Land Data System: Report of the Program Planning Workshops. Pilot Land Data System Working Group, Office of Space Science and Applications. National Aeronautics and Space Administration Technical Memorandum 86250, 1984.

[NASA, 1984b] Earth Observing System. Report of the Science and Mission Requirements Working Group for the Earth Observing System. National Aeronautics and Space Administration Technical Memorandum 86129, 1984.

[NASA, 1984c] Trends in Planetary Data Analysis. Executive Summary of the Planetary Data Workshop, Nov. 29-Dec. 1, 1983. National Aeronautics and Space Administration Conference Publication 2333, 1984.

[NASA, 1985] Space Station Summer Study Report. Report of the Space and Earth Science Advisory Committee Task Force on Scientific Uses of Space Station. National Aeronautics and Space Administration, 1985.

[NASA, 1986] Earth Observing System Data and Information System. Report of the Eos Data Panel. National Aeronautics and Space Administration Technical Memorandum 87777, 1986.

[NRC, 1978] Space Plasma Physics: The Study of Solar-System Plasmas. Volume 1, Reports of the Study Committee and Advocacy Panels. Space Science Board, National Research Council. National Academy of Sciences, 1978.

[NRC, 1982] Data Management and Computation, Volume 1: Issues and Recommendations, Report of the Committee on Data Management and Computation, Space Science Board, National Research Council. National Academy Press, 1982.

[NRC, 1984] Solar-Terrestrial Data Access, Distribution, and Archiving. Report of the Joint Data Panel, Committee of Solar and Space Physics, Space Science Board and of the Committee on Solar-Terrestrial Research, Board on Atmospheric Science and Climate, National Research Council. National Academy Press, 1984.

[NRC, 1986] Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences. Committee on Data Management and Computation, Space Science Board, Commission on Physical Sciences, Mathematics, and Resources, National Research Council. National Academy Press. In press.

APPENDIX A

SUMMARIES OF SELECTED STUDY REPORTS

The discussions, conclusions, and recommendations contained in a number of reports of earlier data management studies provided much of the basis for the current study. The published reports (and one which is nearing publication) which were drawn on extensively are cited in the text of the report and are included in the list of references.

Several other reports pertain to the issues under discussion and influenced the course and tone of the study, but are not readily available in the open literature. As a convenience to the reader, summaries of those reports are contained in this Appendix. These summaries highlight the portions of the reports which relate most directly to the current study. For more information about the original content or context, the reader is referred to the full reports.

LIST OF SUMMARIES

MORE EMPHASIS NEEDED ON DATA ANALYSIS PHASE OF SPACE SCIENCE PROGRAMS

REFERENCE

More Emphasis Needed On Data Analysis Phase Of Space Science Programs, National Aeronautics and Space Administration. Report to the Congress by the Comptroller General of the United States. Report number PSAD-77-114, June 27, 1977.

INTRODUCTION

In the transmittal letter the Comptroller General stated that "this report discusses the National Aeronautics and Space Administration's support of investigators' postlaunch data analysis efforts on space science experiments, and improvements needed in making this data available to other members of the scientific community for further analysis. This review is a follow-on to a survey in which we found that data on a number of successfully launched experiments had not been submitted to the National Space Science Data Center as required."

RECOMMENDATIONS

The report's first recommendation is to the Congress:

When evaluating the National Aeronautics and Space Administration's (NASA's) program content and budget requests, the Congress should examine the adequacy of NASA's allocation of resources between gathering space science data and analyzing it. Greater emphasis is needed during the data analysis phase of a program to obtain the maximum scientific benefit from the data obtained.

The other recommendations are that the Administrator of NASA should:

1. Direct the Associate Administrator for Space Science to enforce the contracts and in-house agreements requiring investigators to submit data to the National Space Science Data Center.

2. Direct the Associate Administrator for Space Science to maintain a schedule showing when investigators are expected to submit data from their experiments, and to set up a system showing which experiments should receive priority attention at the National Space Science Data Center.

3. Develop more realistic estimates of funds and time necessary to adequately support investigators' data analysis.

4. Assign certain data acquisition duties to project scientists.

The elaboration in the report on item 4 expressed the belief that an alternative to hiring additional acquisition scientists for the NSSDC is to expand the roles and responsibilities of NASA's space science project scientists to include the NSSDC's data acquisition responsibilities. It uses as the basis for that recommendation the outline of the roles and functions of the project scientist contained in attachment E to NASA Management Instruction 7100.11 dated June 20, 1975.

GEOPHYSICAL DATA INTERCHANGE ASSESSMENT - 1978

REFERENCE

Geophysical Data Interchange Assessment - 1978. Report of the Committee on Data Interchange and Data Centers, Geophysical Research Board, Assembly of Mathematical and Physical Sciences, National Research Council, National Academy of Sciences. 1979

INTRODUCTION

In March 1977 the Committee on Data Interchange and Data Centers, with the approval of the Geophysics Research Board, established *ad hoc* panels in six geophysical disciplines to study the World Data Center - A's and their associated National Data Centers. The individual panel reports are included as appendices to the committee report. The main text of the committee report represents its summarization and integration of the panels' contributions.

The membership of the committee at the time of the report consisted of:

Thomas O. Haig (Chairman)
Richard Y. Dow (Executive Secretary)
Sidney A. Bowhill
Colin B. Bull
Carl W. Kreitzberg
Juan G. Roederer
Bruce A. Taft
M. Nafi Toksoz

STUDY OBJECTIVE

The objective of the Committee on Data Interchange and Data Centers and its contributing panels was to understand better the dynamics of present National Data Center (NDC) and World Data Center - A (WDC-A) activities and to recommend ways to improve data services. The report stated that in many cases their recommendations were intended to reinforce actions already started by one or more of the responsible agencies.

This report focuses exclusively on data center activities. It does not treat related data utilization issues such as the provision of project data to the investigators, the investigators' analysis activities, and collaborative activities between investigators which do not involve the data centers.

FINDINGS AND RECOMMENDATIONS

The consolidated findings and recommendations are listed in succinct form in Chapter 1 of the report. All of them are of interest in the context of the current study, and are quoted as follows:

"1. In the United States, the growth of concern about the environment is producing a corresponding increase in the flow of geophysical date. To accommodate this growth, we recommend continued expansion and automation of data-center facilities.

"2. The scientific community will become more heavily dependent on effective data-center services in the future; thus, we recommend that this community lend its strong support to the achievement of an adequate level of funding for those services.

"3. In planning data-intensive projects, we recommend that the funds necessary for preparation of data for archiving and for long-term preservation and distribution be included from the outset.

"4. We recommend that the application of modern techniques to make the data available to users be accelerated. When necessary, special action should be taken to accelerate automatic-data-processing (ADP) procurement procedures.

"5. Some geophysical observations made regularly by federal agencies and scientists are not deposited in National Data Centers (NDC's). Where current or anticipated circumstances could result in the loss of unique and valuable data sets, we recommend that the Data Centers be provided the resources necessary to take custody of such data.

"6. We recommend that the National Aeronautics and Space Administration (NASA) re-evaluate the arrangements for processing and archiving space data in the National Space Science Data Center (NSSDC), so that adequate capability is assured to meet NASA's responsibilities to the user community.

"7. We recommend that scientific use of the data be promoted by encouraging and financing scientists and students, perhaps through a program of data scholarships, and, in some cases, by stimulating scientific activity by the Data Center personnel.

"8. We recommend expansion of contacts with WDC's in other countries to reach agreement on common formats for related data, to ensure more timely data exchange, to expand the types of data that are exchanged, and to explore the possibility of digital data links.

"9. We recommend that the Geophysics Research Board consider the need for regional or specialized data centers in such subjects as mesoscale climatology, water resources, and snow and ice data.

"10. Because most records at the Data Centers would be difficult or impossible to replace, and most of the present archives are inadequately protected against loss through natural or human causes, we recommend that duplicate copies be made of all physical data and archived in a secure place.

"11. In view of the importance of the <u>Guide to International Data Exchange Through the World Data Centers</u>,
we urge the International Council of Scientific Unions Panel on WDC's to issue a new and revised <u>Guide</u> promptly.

"12. Believing that there is need for a unified national geophysical data policy, the Committee on Data Interchange and Data Centers proposes to assemble a draft policy during the next several months. Input by interested government agencies, National Research Council committees, professional societies, and other organizations and individuals will be solicited as part of this process."

OTHER POINTS OF INTEREST

This report makes a very useful distinction between three classes of data, (1) data from operational and quasi-operational activities, where the collection is reduced to a routine, and the data are provided to a data center as though it is simply another among many users, (2) data collected through individual experimental programs, where there must be a substantial extra effort by the experimenters to place their data sets in the data centers, and (3) what it calls "private data sets". These are data sets which may be developed over a long period of time in the course of an investigator's work where the investigator does not have a contractual obligation to place the data in the data center, but where the data may be of substantial value to other experimenters.

The report then points out that "The general philosophy of data center operation places the responsibility for preparing data and accompanying documentation for archival purposes with the agency making the observation, not with the data centers." It observes that this policy appears to operate well and is probably the most efficient alternative when the data collection is a routine function. However, difficulties are experienced where the data are collected as part of an experimental program. And the situation is even worse for the "private data sets" which seldom find their way into the data centers.

The report quotes NASA's long-standing policy that data being collected in experimental programs should be available to the principal investigator first and, after a reasonable length of time, be placed in the National

Space Science Data Center (NSSDC). It observes that the responsibility for placing the data in the NSSDC rests with the principal investigators and the cognizant NASA Project Offices. It further observes that frequently the principal investigators has had neither the interest nor the funds to process the data for archiving, and the NASA Project Offices had frequently ceased to exist once the data collection phase was complete. A consequence of this was that the data were often received by the NSSDC with inadequate documentation and in incomplete or otherwise unusable form. By the mid-70's the NSSDC staff had developed the capability to reprocess the data to make them generally useful, but that was stopped in the spring of 1977 when NASA management re-established the original policy of principal investigator and Project Office responsibility for this function. At the same time they removed the resources from the NSSDC which permitted them to perform this function. However, no fundamental changes were made by the Project Offices, so that this problem persists.

The committee suggested that "a mechanism should be created to search for and to identify those private data sets that are of sufficient value to other data users to justify the cost of documenting and formatting the data for inclusion in the NDC's". It further suggested that "Documentation and processing of the private data sets for archiving at the appropriate center probably could be done best by those currently holding the data sets."

Relative to data access, the committee observed that catalogs of data held by the centers are essential if the data are to be generally useful. These should be available in printed and, in some cases where holdings are large, magnetic-tape forms. Furthermore, computer-to-computer access to information on data availability and processing history and to the data sets themselves should be implemented to serve certain classes of users. It pointed out that correlative studies (in which relationships between different parameters are sought) generally require data from various measurements and disciplines. It further emphasized that communication and direct access between data centers would reduce the cost and time required to perform these correlative studies.

The committee expressed the view that "regional centers" (sometimes referred to as "discipline centers") could provide improved user services and relieve the national centers of a significant portion of their data service requests. It suggested in the report that the need for such centers devoted to mesoscale climatology, regional water resources, seismology, and regional snow and ice data be considered by the appropriate committees of the National Research Council.

NATIONAL SPACE SCIENCE DATA CENTER
REVIEW - 1982

REFERENCE

J. H. Trainor. Summary of Notes and General or Consensus Conclusions. Report submitted to the Assistant Associate Administrator for Space Science and Applications, National Aeronautics and Space Administration, July 1982.

INTRODUCTION

In July 1982 the National Space Science Data Center (NSSDC) Review Committee was established under the authority of the Assistant Associate Administrator for the Office of Space Science and Applications. The committee did most of its work at a two day meeting at the Goddard Space Flight Center on 12-13 July 1982. The referenced report was prepared by Dr. J. H. Trainor and concurred in by the committee members. A presentation of the results of the review were made to the Office of Space Science and Applications in NASA headquarters on 3 August 1982. The report was transmitted to the Assistant Associate Administrator for Space Science and Applications on 8 September 1982.

The membership of the NSSDC Review Committee consisted of:

James Trainor (Co-chairman)	Goddard Space Flight Center
Tony Villasenor (Co-Chairman)	NASA Headquarters
Ray Arvidson	Washington University
Peter Banks	Stanford University
Sara Heap	Goddard Space Flight Center
Chris Russell	University of California, LA
Ethan Schrier	Space Telescope Institute
John Simpson	University of Chicago
Verner Soumi	University of Wisconsin
Don York	Princeton University

COMMITTEE CHARTER

The NSSDC Review Committee was established for the purpose of reviewing the NSSDC position with respect to archiving and accessing space science data. Specific charges to the committee were to:

1. Assess the performance of the NSSDC to date.

2. Review the funding posture of the NSSDC, assess its adequacy and distribution, and make recommendations as appropriate.

3. In the light of current and forecasted science needs and technology, recommend future technical and programmatic directions for the NSSDC.

It should be noted that the purpose of the review was to look at the NSSDC as an institution rather than to look at the "NASA data system". An immediate output was desired to assist NASA Headquarters in optimizing funds and resources for the NSSDC.

SUMMARY OF CONCLUSIONS

The main body of the report was a summary of the notes taken during the discussions by Dr. Trainor and his attempt to reduce the discussions to a consensus position for the committee. Since this material is brief in the original report it is reproduced here in its entirety.

> "1. There is a consensus that the NSSDC has served some, but not all of its functions generally well over the past 10 years; but, now technology is overtaking it. The NSSDC has done rather well for the magnetospheric and interplanetary data, less so for astronomical data and not well for digital planetary imaging data. (This latter point is controversial in the sense that the NSSDC says JPL did not want to give up the data and JPL says the NSSDC did not want it!)
>
> "2. Even the users who were relatively happy with the service and product of the NSSDC were quite unhappy with the routine response times which run from some weeks to several months.
>
> "3. Many users of the NSSDC are apparently searching for ancillary data or correlative data and specifically request the data in the form of plots, microfilm and/or microfische and specifically not as magnetic tapes. Requests to the P.I.'s in the early stages of a mission reflect this also. Primary data users (fields and plasmas detailed correlation, power spectrum analyses, and planetary images for instance) need extensive digital data. The latter users are very persuasive that the proper way to handle their needs is via on-line data access e.g., a data base distribution system which does not exist within NASA.
>
> "4. While the NSSDC has instituted semiannual review and collection procedures to ensure that Principal Investigators and Science Teams do put data in the NSSDC, the system overall is weak. Percentage collected is much better over the last 3 years than before, but it is a fact that when data analysis funds are short, the quality as well as the quantity of the delivered product suffers. It seems that the funds for each P.I. or team for his/her NSSDC data effort should be a separately budgeted and funded item in their contracts, and that the contract should specify the data form, quality, accomplished processing and delivery dates, for instance. Reference was made to the current HEAO (MSFC) effort in this regard. The NMI 8030.3A concerning data obtained

from space science flight investigations is general and weak. If it were strong and specific, the NSSDC would be better directed and served. The committee recommends that CODMAC or this committee be directed to write a draft of such a NMI.

"5. It is clear that the NSSDC has been pushed into a position over the years of a low-cost-and-proud-of-it system. The civil service staffing of the NSSDC is below anybody's minimum in both quantity and general technical level, especially as regards new technology applied to the archival and data distribution function. The NSSDC has a very cost effective contractor, but again they largely carry out the day-to-day functions and generally do not carry out innovative high technology work. Current budgets do not allow it.

"6. It is apparent that the Coordinated Data Analysis Workshops (CDAW's) run by the NSSDC are a high point of its operation. They work well for the magnetospheric groups and interplanetary physics groups which have been involved. It was the feeling of the committee that they worked especially well because there were interested discipline scientists at the GSFC and at the NSSDC who ensured their success. Also the NSSDC data is dominated by "fields and particles" data. There were discussions with mixed results as to whether such CDAW's should be funded in a basic NSSDC submission or that each responsible Program Office should fund CDAW's in their own area separately.

"7. There is a general need to get competent technical people into support roles in the NSSDC. The jobs must be interesting and attractive in order to keep an innovative, able staff. It was thought in the committee that a good function for a national facility would be to provide the lead in advanced technology in scientific data archiving.

"8. In addition to the comments relative to the NASA staff, the committee suggested the active involvement of an intern program, sabbaticals at the NSSDC, etc., for the purpose of stimulating and informing the NSSDC and the users of the latest needs, abilities, innovations, etc.

"9. There is a need for discipline subgroups to advise the NSSDC on how to archive data, what form the data should take, what to save, transfer, distribution capability, etc.

"10. There is a need for a complete on-line data catalog. A catalog that is on-line for other facilities/computers to browse through. Especially true for astronomical and imaging data generally is a descriptor or set of notes explaining the circumstances of the observation, gains, filters, exposure, observer if any, etc.

"11. There is some unhappiness with the simple but arbitrary charge policy for the NSSDC user services. Most users are charged nothing, but currently if the NSSDC staff feel that a request is abusive, a charge is made. A phone call should be made to the requester informing him of the charge and generally about the amount of data he/she could receive in what period at no charge.

"12. There is a concern regarding the lifetime of data on magnetic tape, and the absence of an active plan to demonstrate such lifetime.

"13. And finally, the committee generally was unhappy that their charter did not allow them to address the larger question of data transfer/technology/distributed data sets.

ADDITIONAL COMMENT

In his memorandum transmitting the report to the Assistant Associate Administrator for Space Science and Applications, Michael Devirian stated, "During the course of the review, it became clear that an objective evaluation of the NSSDC was made difficult by the lack of an overall coherent NASA posture on data archiving. It is strongly recommended that this, or a similar committee, be chartered to provide recommendations from the science community regarding a comprehensive archive policy, and that these recommendations be used as a basis for archive planning by an appropriate Agency group."

RECOMMENDATIONS OF THE DATA SYSTEMS USERS WORKING GROUP

REFERENCE

Recommendations of the Data Systems Users Working Group. NASA. August 1983.

INTRODUCTION

The Solar Terrestrial Division of the NASA Office of Space Science started the Space Plasma Computer Analysis Network (SCAN) in May 1980. Its motivation was the realization that most new work in space plasma physics requires data from a diversity of sensors in a variety of locations in space, and involves the collaborative efforts of many investigators at numerous locations to correlate and analyze the data. It was believed that this work could be facilitated at minimum cost by the creation of such a network.

It was also realized from the beginning that the concept could be made to work well only by the inclusion of the researcher user community in the planning, building, and operation of the network. Thus, a Data System Users Working Group was established in September 1980 to provide a mechanism for interaction between the system builders and the research community. This working group has been continuously active to provide guidance to the Solar Terrestrial Division (now the Space Plasma Physics Branch in the Earth Science and Applications Division) in the Office of Space Science and Applications).

This SCAN activity has been so successful that it is now considered by many to be a good basis for expanding the concept into other research areas where the same need for correlative studies by the use of pooled resources exists. This expansion is occurring in small steps as resources can be found, under the name Space Physics Analysis Network (SPAN). The Data Systems Users Working Group is remaining active in this broadening of the concept. Their latest recommendations reflect their experience in establishing the SCAN, and are so pertinent to this study that their short paper is included here in its entirety.

The members of the Data Systems Users Working Group in mid-1982 were as follows:

Policy Management Subgroup:

Rod Heelis (Chairman)	University of Texas, Dallas
Ron Zwickl (Co-chairman)	Los Alamos Nuclear Laboratory
Vincent Abreu	University of Michigan
Joe Barfield	Southwest Research Institute

Rick Chappell	Marshall Space Flight Center
Andy Christensen	Aerospace Corporation
Tom Cravens	University of Michigan
Eugene Greenstadt	TRW, Inc.
Elaine Hansen	LASP, University of Colorado
Dave Klumpar	University of Texas, Dallas
Bob McPherron	University of California, LA
Richard McEntire	Applied Physics Laboratory, JHU
Pat Reiff	Rice University
Bob Theis	Goddard Space Flight Center
Tony Villasenor	NASA Headquarters
Ray Walker	University of California, LA
Mike Wiskerchen	NASA Headquarters

Software and Hardware Subgroup:

Joe Doupnik (Chairman)	Utah State University
Bill Peterson (Co-chairman)	Lockheed Aerospace Corporation
Joe Bredekamp	Goddard Space Flight Center
Dennis Gallagher	University of Alabama
Ron Janetzke	Southwest Research Institute
Barry Mauk	Applied Physics Laboratory, JHU
Carl McIlwain	University of California, SD
Doug Menietti	Southwest Research Institute
Tom Miller	Standard Oil Company
Thomas Moore	University of New Hampshire
Lee A. Reinleitner	University of Iowa
Don Sawyer	Goddard Space Flight Center
Howard Singer	Air Force Geophysics Laboratory
Lora L. Suther	Applied Physics Laboratory, JHU
Bill Taylor	TRW, Inc.
Roy Torbert	University of California, SD

Data Base Standards Subgroup:

Charles Sonett (Chairman)	University of Arizona
Randy Davis (Co-chairman)	LASP, University of Colorado
Ted Fritz	Applied Physics Laboratory, JHU
David Kayser	Aerospace Corporation
Richard Munro	High Altitude Observatory, NCAR
Bob Power	University of Texas
Jim Slavin	Jet Propulsion Laboratory, CIT
Paul Smith	Goddard Space Flight Center
Ted Tarbell	Lockheed Aerospace Corporation
Bruce Tsurutani	Jet Propulsion Laboratory, CIT
Jim Vette	Goddard Space Flight Center

Networking and Standardization Subgroup:

Rob Gold (Chairman)	Applied Physics Laboratory, JHU
Jim Green (Co-chairman)	Marshall Space Flight Center, NASA
Bob Clauer	Stanford University
Neal Cline	University of California, LA

Jan Hauser	NASA Headquarters
Dick des Jardins	Computer Technology Associates
Julian Johnson	Marshall Space Flight Center
Douglas Potter	University of Washington
Bob Stevens	NASA Headquarters
Hunter Waite	Marshall Space Flight Center
Jim Willet	Jet Propulsion Laboratory, CIT
Fred Wulff	NASA Headquarters

COMPLETE TEXT OF THE PAPER

"Introduction. A considerable evolution has occurred in the past two decades in the disciplines of Solar-Terrestrial and Interplanetary Physics. Early research was centered around exploratory missions in which measurements from individual scientific instruments could be meaningfully employed to advance our state of knowledge. As these scientific disciplines have advanced, a much more profound, and interrelated, set of questions is being posed by researchers. The result is that present-day investigations are generally much more complex: large volumes of data are acquired from multiple sensors on individual spacecraft or from ground based systems and, quite often, data are needed from multiple sources in order to address particular physical problems.

"It is clear that research in solar-terrestrial physics during the 1980's, and beyond, will be devoted to intense multidisciplinary studies aimed at exploring very complex physical questions (cf. Solar-Terrestrial Research for the 80's, NAS). It is in this spirit that the Data Systems Users Working Group (founded in 1980 and currently composed of representatives of over forty university, industrial, and government institutions) recognizes that major future advances in solar and space physics will require close collaboration among investigators through interactive exchange of scientific information. Increasingly, scientists spend large amounts of time contacting other researchers to obtain data needed to solve given problems. Long time delays in such efforts typically result due to other commitments of collaborating scientists. The problems of data exchange are exacerbated by the lack of standards for scientific data bases. Each spacecraft project, for example, designs a new data system within the context of the project's immediate requirements, with very little thought given to the data system's scientific usability and its compatibility with existing data bases. The net result is that, at present, most researchers recognize the value of multidisciplinary studies, but the cost in time and effort is devastating to the research efforts. This trend is antithetical to the needs of solar and space physics research.

"Recommendations. The DSUWG recognizes that computer networking holds the most promise of meeting collaborative scientific requirements in the most efficient and cost effective manner for archived, current, and future data bases. The DSUWG therefore recommends that NASA:

> "Establish a solar and space physics pilot program to create a Space Physics Analysis Network (SPAN) that would link together a large number of NASA space scientists.

"Further, the DSUWG recommends that SPAN should conform to the following guidelines:

> "1. The network must function such that users' needs are addressed. This could be insured by designating the DSUWG as an advisory committee to the project manager and project scientist guiding the development of SPAN.
>
> "2. The DSUWG will carry out continuous review and guidance of the network.
>
> "3. SPAN should maintain a stable environment for facilitating correlative scientific research with the flexibility of being a test bed for the development of improved technologies.
>
> "4. SPAN should be built by using available, but state-of-the-art, components (from hardware through software) including standard data and graphics formats which are upwardly compatible.
>
> "5. SPAN should become a test bed for the design of data systems for future projects.
>
> "6. The inclusion of the NSSDC within this network to act as a central library and data catalog center is highly desirable at a very early stage. This same recommendation has been made by the CSSP/CSTR data panel.
>
> "7. SPAN should use as its foundation the current SCAN system, based at MSFC. However, the architecture of the network should be defined so that some redundancy exists. Since NASA Code T is involved presently in paying communication costs for SCAN, it is essential that coordination with Code T continue throughout the pilot program. In addition, the MSFC node would benefit from continued interaction with Code R.
>
> "8. MSFC should be identified initially as the lead NASA field center in the pilot program. Based upon present extensive efforts under way there.
>
> "9. The solar and space physics pilot should coordinate its efforts with those of other pilot programs within NASA, and with other interested agencies (e.g. NOAA, DOE).

"<u>Further Considerations</u>. Computer networking is a concept which is presently employed in various capacities throughout the world. However, the application of networking to various fields of science, with a few exceptions, is lacking. The DSUWG feels that the creation of a computer network which serves the needs of the solar and space physics community could have many productive benefits:

"1. The proposed SPAN network would facilitate current correlative research by rapid data exchange.

"2. The network would reduce the time and effort spent in locating and acquiring diverse data sets.

"3. Costly software development time to convert acquired data would be reduced since the data would arrive in a directly usable form.

"4. The adoption of standard data formats and standard graphics files would further reduce software costs -- the single most expensive aspect of computing.

"5. Such standards would improve portability and allow visiting scientists to be more productive by requiring much less time to learn to operate in the new environment with its new data formats and data systems.

"6. The ability to obtain usable software developed at a remote site would increase.

"Rapid access to required data allows the scientist to test his ideas while they are still fresh in his mind. The availability of data also allows the researcher to analyze the problem in more detail within a reasonable time frame. As the network grows, we envision that the individual scientist will gain greater flexibility. As an example, there would be easier access to supercomputers and other scarce resources and regional data centers could also be brought on line. We further recognize that outside agencies (e.g., NOAA and DOE) could be included within SPAN at their own cost. Of particular importance is the fact that new NASA projects could be integrated into the existing network, thereby insuring a common direction and future compatibility."

STUDY BY THE PANEL ON GEOPHYSICAL DATA AND PUBLIC POLICY

INTRODUCTION

A Panel on Geophysical Data and Public Policy was established by the Geophysics Study Committee (under the Geophysics Research Board, Commission of Physical Science, Mathematics, and Resources, National Research Council, National Academy of Sciences). The emphasis indicated by the panel's name resulted from the realization that geophysical data collection and utilization is no longer a purely scientific matter, but has broader societal implications. The charge to the panel was to carry out a study that would:

1. Determine the scope of the problems associated with geophysical data.

2. Establish in terms persuasive to the scientific community and to those concerned with generation and management of data why there should be a national geophysical data policy.

3. Lay out a plan to develop such a policy.

The membership of the panel in 1983 was:

Michael A. Chinnery (Chmn.)	NESDIS, NOAA
Shelton Alexander	Pennsylvania State University
Arthur G. Anderson	IBM, Inc.
Johathon A. Brownell	Dartmouth College
Martin L. Ernst	Arthur D. Little, Inc.
Jerome A. Eyer	Grace Petroleum Corporation
John W. Firor	National Center for Atmospheric Research
Thomas O. Haig	Black Earth, Wisconsin
William J. Hinze	Purdue University
Allan V. Kneese	Resources for the Future
James F. Lander	National Geophysical Data Center, NOAA
David W. Moody	U.S. Geological Survey, USDI
G. Wesley Rice	CONOCO, Inc.
Robert L. Rioux	Minerals Management Service
Juan G. Roederer	University of Alaska
Carl H. Savit	Western Geophysical Corp. of America
Erwin R. Schmerling	NASA Headquarters
Alan H. Shapley	NOAA
George Swanlund	Control Data Corporation
Gilbert F. White	University of Colorado
Staff support:	
Thomas M. Usselman	National Academy of Sciences

DISCUSSION

The panel has found the carrying out of its charge to be frustratingly complex. There appears to be little consensus on which of the problems associated with geophysical data are the most important, or on their solutions. The panel is still debating these questions; therefore it is premature to attempt a listing of their probable conclusions and recommendations. There are, however, several consistent threads which are emerging from its work and which are important to this study.

The panel has discussed a set of problems which fall under three general headings, scientific, data management, and policy. The scientific problems center around the fact that the scientists are very possessive of the data collected in the process of their investigations while they are analyzing them, but become disinterested in the ultimate disposition of the data once their use of them is completed. Thus, it is difficult to obtain the absolutely essential involvement of the scientists in the processes of deciding which data to archive, preparing the data for archiving (including quality control), and determining which data should be purged from the archives.

The data management problems include the lack of useful compilations of comprehensive data inventories and of referral systems, the inability thus far to make full use of state-of-the-art technology because of the lack of system planning and funding, and the difficulties in devising an integrated data management strategy when the participants cannot even agree on whether the data holdings should be collected in a Data Center or in a less centralized, distributed system.

And the policy problems include the lack of overall policy guidance to help unify the different agency approaches, difficulties in budgeting and allocating for data storage and dissemination, and the difficulties in coping with proprietary or classified data.

The panel seems to be in general agreement that scientists need to develop a new sense of responsibility for the data that they collect, including making them available for others to use. There is a feeling by the panel that, in the areas of geophysical research where the cost of acquiring the data is so high, the scientists need to realize that the data on which they base their papers may have an overall value which exceeds the value of their individual papers. Thus, data management can not be treated as an afterthought, a bit of burdensome house-cleaning to be done once the project is over, but should be an integral part of the entire project, from initial planning to completion.

APPENDIX B

NASA POLICY DOCUMENTS RELATED TO SPACE RESEARCH DATA MANAGEMENT

A number of NASA directives bear on space research data management. This Appendix lists the most directly applicable. The documents are presented in numerical order. NMI indicates a NASA Management Instruction, NHB a NASA Handbook, NPD a NASA Policy Directive, and HQMI a Headquarters Management Instruction.

NMI 1440.6B	NASA Records Management Program, Dec. 14, 1979
NHB 1441.1A	NASA Records Disposition Handbook, Dec. 1, 1970, with changes 1, 2, and 3 dated Aug. 14, 1972; Feb. 1973; and Feb. 1974.
NMI 2220.5B	NASA Scientific and Technical Information, Jul. 12, 1983
NMI 2410.6	NASA Software Management Requirements for Flight Projects, Feb. 1, 1979
NMI 7100.11	Announcement of Opportunity Process-Acquisition and Administration of Space Science Investigations, Jun. 20, 1975
NMI 7121.1C	Planning and Approval of Major Research and Development Projects, Mar. 24, 1977
NPD 8000.1	NASA Earth Resources Survey Program Imagery, May 10, 1972
NMI 8030.3A	Policy Concerning Data Obtained from Space Science Flight Investigations, May 2, 1978
NHB 8030.6A	Guidelines for Acquisition of Investigations, Nov. 28, 1978
NMI 8320.1B	Basic Policy for NASA University Relationships, Aug. 1, 1978
NMI 9080.1B	Review, Approval, and Imposition of User Charges, Oct. 19, 1978

APPENDIX C

NASA MANAGEMENT INSTRUCTION 8030.3A

18646

[7510-01]
Title 14 - Aeronautics and Space

CHAPTER V - NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

PART 1205 - SPACE SCIENCE FLIGHT INVESTIGATIONS

Subpart 1205.1 - Policy Concerning Data Obtained From Space Science Flight Investigations

AGENCY: National Aeronautics and Space Administration.

ACTION: Final regulations.

SUMMARY: The National Aeronautics and Space Administration (NASA) finalizes revision to its policy concerning data obtained from Space Science Flight Investigations. This revision is necessary to redefine the roles and responsibilities of individuals conducting space science investigations, Federal officials concerned with implementation of programs, and the National Space Science Data Center (NSSDC) to insure proper use and preservation of space science data. The revised policy, through the use of a Project Data Management Plan, will reduce the amount of data acquired by the NSSDC and insure that the data deposited in the NSSDC is processed in the most efficient, cost effective manner.

DATE: Effective date May 2, 1978.

ADDRESS: SRT Coordinator, Office of Space Science, National Aeronautics and Space Administration, Washington, D.C. 20546.
FOR FURTHER INFORMATION CONTACT:
Franklin G. Tate, 202-755-3833.

SUPPLEMENTARY INFORMATION: On December 28, 1977, a notice of revision to the Policy Concerning Data Obtained from Space Science Flight Investigations was published in the **Federal Register** (42 FR 64706). The purpose of the revision is to redefine responsibilities of individuals concerned with implementation of programs and the NSSDC to insure proper use and preservation of space science data in the most cost efficient manner. Interested parties were granted 30 days to submit written comments regarding the proposed revisions. No written comments were received.

The proposed revised Policy Concerning Data Obtained from Space Science regulations is hereby adopted without change, as set forth below.

ROBERT A. FROSCH,
Administrator.

Subpart 1205.1 - Policy Concerning Data Obtained From Space Science Flight Investigations

Authority: 42 U.S.C. 2454

Subpart 1205.1 - Policy Concerning Data Obtained From Space Science Flight Investigations

1205.100 Scope.

(a) This subpart continues the policy and responsibilities for reduction, analysis, deposition, preservation, and dissemination of data obtained from space science flight investigations. For glossary of terms and abbreviations, see Appendix A to this subpart.

(b) These provisions apply to all data obtained from space flight experiments processed by the Space Science Steering Committee and approved by the Associate Administrator for Space Science. The policy and procedures for the conduct of the Space Science Program and the responsibilities for the selection and support of scientific investigations and Investigators are set forth in NASA Management Instruction 7100.11 and NASA Handbook 8030.6.

1205.101 Policy.

(a) In conducting space science flight experiments, NASA shall seek to:
(1) Preserve the integrity of each investigation.
(2) Encourage the participation of the best qualified scientists.
(3) Make the results of investigations generally available to the scientific community at the earliest practicable time.
(4) Document the instrument performance and probable cause of malfunctions that occur.

(b) NASA shall rely on individual scientists as Principal Investigators (PIs) to carry out a complete investigation by:
(1) Selecting, when appropriate, associates known as Co-Investigators (COIs) who have supporting roles in the investigation.
(2) Conceiving specific investigations.
(3) Developing, when appropriate, the instrumentation for the investigation.

(4) Participating actively, wherever appropriate, in the actual conduct of the investigation.
(5) Reducing and analyzing the data obtained.
(6) Publishing their findings as soon as practicable.
(7) Making their Reduced Data Records (RDRs) and Analyzed Data Records (ADRs) along with supporting documentation available on a timely basis for use by the scientific community and the news media, in accordance with a Project Data Management Plan (PDMP).
(8) Documenting any significant malfunctions which occur during the lifetime of the experiment.

(c) For certain missions, NASA shall also rely on Guest Investigators to obtain data within the capability of a given mission, which are additional to the mission's primary objectives, and to perform an analysis of the acquired data. Such projects generally maintain a data retrieval and dissemination system or provide a data analysis system for the life of the project. In such cases the PDMP should address the exchange of information with the National Space Science Data Center (NSSDC) so it can act as a switching center by referring data requesters to the appropriate contact or making requests on their behalf and handling the distribution. In addition, the eventual transfer of appropriate data to a more permanent repository prior to termination of the project shall be addressed in the PDMP.

(d) For facility-class payload missions, NASA may rely on Guest Investigators to obtain and analyze data. In connection with such missions a complete analysis activity may be provided for the Investigators; this activity may involve the creation of an institute with a significant lifetime. In such cases the PDMP should address the exchange of information with the NSSDC so it can act as a switching center by referring data requesters to the appropriate contact or making requests on their behalf and handling the distribution. In addition, the eventual transfer of appropriate data to a more permanent repository prior to termination of the mission activity or institute shall be addressed in the PDMP.

(e) A provision for the release of data obtained by the individual Investigator from the investigation shall be included in an agreement with the Investigator at the time of selection to participate. NASA shall take such action as necessary to insure that data are released as required to meet scientific, technological, and public information needs.

(f) Foreign scientists participating in cooperative space science flight investigations shall be governed by appropriate international agreements and/or memoranda of understanding.

1205.102 Responsibility.

(a) OSS. The Associate Administrator for Space Science is responsible for the issuance of implementing management instructions and guidelines consistent with the provisions of this subpart. The Assistant Associate Administrator for Space Science (Science) is responsible for the program management of the NSSDC and for insuring that any noncompliance with PDMPs will be factored into future selection considerations.

(b) OSS Program Scientist.

(1) Each Program Scientist is responsible for establishing the data analysis policies for each mission including data sharing and collaborations on data analysis. He/she will review the PDMP to assure that data interpretation meetings will be conducted, that wide dissemination of data through presentations and publications will occur and that data and supporting information will be made available to the scientific community in accordance with the PDMP and, along with the Project Scientist, he/she will monitor the execution of the PDMP. He/she will assist NASA public affairs personnel in meeting public information needs.

(2) Each Program Scientist will insure that the letter of notification of selection stipulates that the Principal Investigator (PI), Team Leader (TL), or Guest Investigator (GI) contributes to a PDMP prior to receipt of flight data which documents the plans for data analysis, dissemination of results and for making ADRs, RDRs and supporting documentation available to the scientific community through the designated data dissemination facility. The notification letter will further stipulate that selection for further opportunities will be jeopardized by failure to meet the commitments of the PDMP. The Program Scientist is also responsible for sending information for the Space Investigations Documentation System (SIDS) to the NSSDC after letters of notification have been sent.

(c) Office of Management Operations. The Head, Scientific and Technical Information Branch is responsible for the issuance of instructions to the Scientific and Technical Information Facility (STIF) to provide to the NSSDC a monthly listing of newly acquired articles and documents that contain information about results from NASA-supported space science flight experiments, especially those which can be identified through the contract, grant, or NASA Unique Project Numbers (UPN) supplied by the NSSDC.

(d) Field installations. NASA field installations assigned project management responsibility for space science flight projects are responsible for:

(1) Insuring that the project plan includes a statement within the project results (or equivalent) section that the Project Scientist or Mission Scientist, the appropriate scientific personnel associated with the investigations that comprise the mission, and the NSSDC acquisition manager will develop a PDMP. This PDMP will be approved by the project manager with concurrence signatures by the Project Scientist and the Director of the NSSDC.

(2) Insuring that the contracts or written agreements negotiated between the PI's, TL's, or GI's institution and the project management center specify the responsibility of the PI, TL, or GI for data reduction, data analysis, publication of results, and, where appropriate, the preparation of selected ADRs, RDRs and necessary documentation for delivery to a data disseminating repository. The contracts or written agreements will stipulate that the contract number or, in the case of a NASA field center Inves-

tigator, that the UPN number appear in all reports or articles. Copies of all reports and preprints shall be sent to the STIF and to the NSSDC. This responsibility will be documented in a PDMP to which the Investigators, the Project Scientists, the NSSDC Acquisition Manager, and the appropriate scientific advisory groups will contribute. The PDMP will commit the Investigator, where appropriate, to supply to a specified repository the following documentation upon submission of the selected ADRs and RDRs:

(i) General information about each data set, such as:

(A) Form of data set - hard copy, magnetic tape, microfiche, microfilm, photographic film, etc.

(B) Quantity of data set - number of units of the form.

(C) External identification for each physical unit of the data set - spacecraft, experiment, Investigator's internal ID.

(D) Time period covered by the data set.

(E) Quantity by which data set is ordered - time, orbit number, spatial coordinate, etc.

(F) Supporting documentation - tape formats, catalog, directory, indexes, User's Guide, etc.

(G) Brief description of the data set (not to exceed 250 words).

(ii) Specific information about each data set, such as:

(A) Magnetic tape - track density, recording density, recording mode, recording parity, make and model number of computer used, number of files, size of physical records, logical record format with specification of each field, etc.

(B) Photographic or microform - type of film, frame and/or reel numbers where supporting data and description are located, index of frames and each reel, assurance that all reels are quality controlled to allow proper duplication, etc.

(C) Hard copy - assurance that copy is clean, legible and of proper contrast so it can be photographed; index or catalog if appropriate; form of binding (burst, unburst, loose leaf, bound); etc.

(iii) A Data Users' Guide which includes a summary of the investigation, a description of the instrument, a discussion of calibration procedures and results, a discussion of pertinent events in the operational history that might affect data interpretation, a discussion of over-all data reduction procedures used in generating the various data sets, and other information useful to a scientifically trained recipient of the data.

(iv) A list of all published articles related to the investigation by the investigator group or team and copies of all reports and preprints.

(3) Insuring that Investigators on these projects fulfill the stipulations of the contracts or written agreements pertaining to the responsibilities described in paragraph (d)(1) of this section.

(4) Delivery of EDRs to Investigators on a timely basis. Conditions for discarding or destroying EDRs shall be specified in the PDMP.

(5) Providing to the NSSDC during the writing of the project plan (and provided updates during revisions and specifically after launch) the following information:

(i) Brief statement of the mission objectives for each launch (not to exceed 200 words).

(ii) The names, addresses and telephone and telex numbers of the program manager, Program Scientist, project manager, and Project Scientists for each mission.

(iii) The launch site, launch vehicle, spacecraft weight, and planned orbit parameters.

(iv) A brief description of the spacecraft, not exceeding 250 words.

(v) For each investigation the names, addresses and telephone and telex number of all the Investigators and the relevant contract or UPN numbers.

(vi) The name of each experiment, its weight, average power, and approximate bit rate.

(vii) A brief description of each experiment, not exceeding 250 words.

(e) The PI, TL, or GI (NASA, Non-NASA, and foreign). At the time an investigation is selected, the PI, TL, or GI will be notified by letter signed by the Associate Administrator for Space Science of his responsibilities for data reduction, prime analysis, and the archiving of appropriate ADRs and RDRs. These responsibilities (subject in the case of foreign scientists to the specifications of the governing international agreement) will include:

(1) Completion of data reduction and prime analysis of the data from his experiment within the period of time agreed upon between the PI, TL, or GI and the Associate Administrator for Space Science.

(2) Publication of the results of his analysis as soon as practicable.

(3) Preparation of selected ADRs and RDRs together with the necessary background information to make them usable by other scientists as specified in the PDMP.

(f) National Space Science Data Center (NSSDC). The Director, Goddard Space Flight Center (GSFC), is responsible for management of the NSSDC, the central data disseminating repository for data obtained from space science flight investigations. For functions and operation of the NSSDC, see Appendix B. The NSSDC Director, appointed by the Director, GSFC, is responsible for:

(1) Implementing the NASA project plan for the operation of the NSSDC.

(2) Recommending through the Director, GSFC, any changes in policies, procedures, and plans for the operation of the NSSDC deemed appropriate to the effective attainment of project objectives.

(3) Preparing budget estimates for operation of the NSSDC.

(4) Recommending fees for the computer and reproduction services performed by the Center to Director, Financial Management Divisions, NASA Headquarters, and obtaining from that office a fee schedule which is consistent with NASA practice.

(5) Based upon information contained in the PDMPs, compiling schedules for transmission of ADRs and RDRs to the NSSDC by investigators on NASA space science flight projects.

(6) Assigning an NSSDC acquisition manager to each flight project to participate in the development of the PDMP and to assure that the plan is carried out on an established schedule concerning the deposition of any data and documentation in the NSSDC.

(7) Providing to the STIF a monthly listing of contract, grant, or UPN numbers for all NASA funded space science flight experiments and related investigations.

(8) Preparing guidelines for the submission to the NSSDC of ADRs and RDRs with documentation from non-NASA missions, and disseminating these guidelines to appropriate individuals and agencies to serve in lieu of a formal PDMP.

(9) Reporting through the Director, GSFC, to the Assistant Associate Administrator for Space Science (Science) semiannually on the data acquisition, request activities and financial status of the NSSDC operations.

(10) Assessing adequacy of the NSSDC facilities and the effectiveness of their utilization, and recommending through the Director, GSFC, the necessary actions to meet future facility requirements.

(11) Maintaining, protecting, and retiring NASA records in the custody of the NSSDC in accordance with the policies and practices of the NASA Records Management Program, NASA Records Disposition Handbook (NASA Handbook 1331.1A) and other pertinent management instructions.

1205.103 Support of research.

The NSSDC will support investigations in space sciences by making available its scientific data and facilities. However, the NSSDC will not provide financial support for such research. The Office of Space Science will entertain proposals for space science research based on data available in the NSSDC.

APPENDIX A - GLOSSARY OF TERMS AND ABBREVIATIONS

Analyzed Data Records (ADRs). Those records which the Investigator designates as the best to display the scientific results of an experiment and provide the physical quantities by applying calibration curves or algorithms to the corrected observed quantities of the Reduced Data Records. The data may be time averaged and may incorporate model-dependent assumptions to obtain the physical quantities. Charts, graphs, table, correlation coefficients, model parameters, photographs, and plots are possible forms of these records.

Co-Investigator (Co-I). An associate of the Principal Investigator (PI) who is assigned a supporting role in the investigation. In addition, some data rights may be assigned to the Co-I by the PI.

Experiment. A term used interchangeably with investigation (the latter is preferred). Activity or effort aimed at the generation of data obtained by measurement of space phenomena or the use of space to observe earth phenomena and the resulting analysis of such data.

Experiment Data Records (EDRs). Those records provided to the Principal Investigator, Team Leader, Guest Investigator, Co-Investigator, or team member containing all the data from the mission required to carry out the investigation specified in the contract or launch agreement. These records may include orbital position, spacecraft attitude, instrument attitude, commands, housekeeping data, ground time, spacecraft time, data from other investigations and other information as agreed upon. The exact form of these records and manner in which these data are provided may vary depending upon the policies, procedures, and capabilities of the project, the payload or mission control centers, the data acquisition network, and any support processing facilities. These records shall be specified in the Project Data Management Plan.

Facility-Class Payload Mission. A mission in which the payload is an instrument or set of instruments which serve as a facility for a large group of Guest Investigators who may be selected at different times throughout the life of the mission to participate. This type of mission may not have Principal Investigators or Team Leaders and all the data collected from such a mission is generally maintained by the project for use by Guest Investigators. Availability of data for the scientific community at large shall be specified at the Project Data Management Plan.

Guest Investigator (GI). Investigator selected to conduct observations and obtain data within the capability of a NASA mission, which are additional to the mission's primary objectives, or for a facility-class payload mission.

Investigation. Activity or effort aimed at the generation of data obtained by measurement of space phenomena or the use of space to observe earth phenomena and the resulting analysis of such data.

Investigator. A participant in an investigation. This term may refer to a Principal Investigator, Co-Investigator, Team Leader, team member, Guest Investigator, or any other member of an investigation group.

Mission. One or more flights within an approved payload project.

Mission Scientist. A scientist from a NASA field center assigned to a Spacelab mission, the Mission Scientist has similar functions as the Project Scientist with the exception of direct responsibility for the development of any experiments.

National Space Science Data Center (NSSDC). The main central repository for selected data and documentation from space science flight missions that serves as a disseminator of this archived data and supporting information to users throughout the international scientific community. The NSSDC, located at Goddard Space Flight Center, serves as a switching center for requesters who desire data still held individually by Principal Investigators (PIs) or Team Leaders (TLs) by providing a description of the spacecraft and experiment and the name, address and telephone number of the PI or TL. For missions involving a Guest Investigator program in association with a PI or TL experiment or involving a facility-class payload the role of the NSSDC shall be specified in the Project Data Management Plan.

Principal Investigator (PI). A person who conceives an investigation and is responsible for carrying it out, reporting its results, and providing appropriately selected data and supporting documentation to the scientific community in accordance with the Project Data Management Plan. The PI chooses his Co-Investigators and assigns them roles and privileges. The PI is the primary point of contact with the project office regarding the investigation.

Program Scientist. A NASA Headquarters official assigned to each mission who has a number of roles and responsibilities defined in NASA Management Instruction 7100.11, Attachment D. The most relevant one for this subpart is the responsibility to establish the data analysis, data dissemination, and data archiving policies for the mission, which will be documented in the Project Data Management Plan.

Project Data Management Plan (PDMP). A plan that addresses the total activity associated with the data acquired by a mission from the delivery of the Experiment Data Records to the Investigators to the delivery of selected reduced and analyzed records along with supporting documentation to a specified repository. The plan should provide the milestones in the data reduction, data interpretation, and resource requirements for these phases. Any planned data interpretation meetings, workshops, or other activities should be identified. The type of data records, data products, and compilations that have been selected in concert by the Investigators, the Project Scientist, the Program Scientist, the NSSDC acquisition manager, and any appropriate scientific advisory personnel for general availability to the international scientific community and for delivery to a disseminating repository, such as the NSSDC, shall be specified. For missions where the data will be maintained for many years by the project, the Principal Investigator handling a Guest Investigator program, or by an institute established by the mission, the eventual transfer of appropriate data to a more permanent archive, such as NSSDC or other repository, shall be specified. Conditions for discarding or destroying the Experiment Data Records shall be specified.

Project Scientist. A scientist from a NASA field center assigned to a project to manage the scientific aspects. The roles and responsibilities of this function are given in NASA Management Instruction 7100.11, Attachment E.

Reduced Data Records (RDRs). Those records prepared from the Experiment Data Records by applying corrections, where applicable, for temperature, voltage, gain change, offsets, dead time, drift and other known instrument changes, as well as eliminating unusable noisy periods and periods of questionable instrument performance. The Reduced Data Record should contain all the basic and supporting measurements obtained from the experiment, such as time, position, attitude, settings of instrument by command, housekeeping data, and other information needed to analyze the data in an independent fashion. Visual data, such as photographs derived from imaging processing techniques, may also be considered as RDRs.

Scientific and Technical Information Facility (STIF). NASA's document and report acquisition and abstracting facility that produces a biweekly abstract journal, STAR, covering the aerospace report literature and a biweekly abstract journal, IAA, covering the published literature in these fields. The facility also produces microfiche copies of the report literature for primary distribution.

Space Science Flight Investigations.
Investigations of natural phenomena of the earth and its environment, the moon, other planets, the sun, interplanetary space, and other celestial objects and regions made from aircraft, balloons, sounding rockets, satellites, probes, and manned spacecraft for the purpose of increasing basic knowledge of these natural phenomena. Biological investigations involving the search for extraterrestrial life are included.

APPENDIX B - FUNCTIONS AND OPERATION OF NATIONAL SPACE SCIENCE DATA CENTER

The National Space Science Data Center (NSSDC) has been established at the Goddard Space Flight Center to provide scientific data and facilities in support of investigations in space science.

(a) *Data to be acquired.* The NSSDC will acquire or accept:

(1) RDRs and ADRs from NASA-sponsored space science flight experiments which are deemed appropriate for secondary distribution and archiving by PDMPs or directly by the Program Scientist. The vast majority of records will be from satellite borne instruments. Any departures from a PDMP shall be approved in writing by the Program Scientist, the Project Manager, and the NSSDC Director.

(2) Unclassified ADRs and RDRs made available from space science flight experiments by other agencies in accordance with interagency agreements providing the data media meet standards for reproduction and the supporting documentation is adequate.

(3) Ground-based correlative data, only when it is considered absolutely necessary for the utilization of data in the NSSDC.

(4) ADRs and RDRs from foreign space science flight experiments made available by international exchange of data through the World Data Centers or by cooperative agreements.

(b) *Data not to be acquired.*

(1) Data obtained from operational observations made for specific applications such as weather forecasting, navigation, communication, tracking and telemetry, medical investigations, and technological investigations which contribute only to the development of space flight hardware will not be acquired by the NSSDC.

(2) Any data from space science flight experiments that is excluded from archiving in the NSSDC by a PDMP.

(3) EDRs (including magnetic tapes, telemetry records, exposed film and meteorite collection panels) will not be acquired by the Center.

(c) *Availability of data.* Data records in the NSSDC will be available to users on the following basis:

(1) To U.S. residents and organizations upon request.

(2) To foreign nationals in accordance with procedures of World Data Center A.

(3) To foreign nationals on the basis of cooperative agreements between NASA and the space agencies of foreign governments or multi-lateral organizations devoted to space research.

(4) To foreign governments on the basis of bilateral intergovernment agreements made by the U.S. on behalf of NASA.

(d) *Preservation of the data.* This should be specifically addressed in the PDMP. In general, data in the NSSDC will be preserved for the longest practicable time consistent with the physical life of the record. Records will be reproduced to extend their storage life only if the record of their past utilization justifies such prolongation. Specific categories of data may be reproduced to extend their storage life regardless of past usage in accordance with international or interagency agreements or with the PDMP.

(e) *Interface with other sources of data from NASA missions.* Certain NASA missions such as those involving national facility payloads or those with Guest Investigator programs may maintain the data within the project for many years. The NSSDC will route requesters of such data to the appropriate facility or make request on their behalf depending on the agreement specified in the PDMP. The distribution of catalogs produced by such a mission shall also be determined by the PDMP.

(f) *NSSDC publications.* The NSSDC will issue or provide publications as necessary to facilitate the use of available data. Publications will include:

(1) *Data users guides.* Provides the data user the experiment information and describes the reduced data available. These will usually be written by the Investigators and supplied to the NSSDC as specified in the PDMP.

(2) *Catalogs of data.* Lists of all data from space science flight experiments available from the NSSDC, issued as needed. The forms in which the data are available will also be indicated, i.e., microfilm, tapes, printouts, etc. Catalogs of data available from projects that maintain their own data base, such as Guest Investigator or facility missions, will also be distributed if the PDMP so specifies.

(3) *Data Announcements.* Announces those data sets which are known to have wide appeal when such sets become available.

(g) *Other services.* The NSSDC will provide technical assistance to data users. In some cases, this may involve the conversion of data records into compatible formats to facilitate correlation of data from various sources.

When facilities are available the NSSDC will provide lecture rooms, study rooms, and office space for use by visiting scientists for research involving the use of available data.

(h) *User charges.*

(1) User charges will, as a matter of policy, be in accordance with the policies set forth in the Bureau of the Budget Circular A-25 and NASA Financial Management Manual 9030. The methods to be used in computing the user charges will be reviewed by the Director, Financial Management Division, NASA Headquarters.

(2) Appropriate fees will be charged for reproduction, computer and dissemination services provided to individual users by the Data Center. The NSSDC may perform conversion of data records and general technical assistance without charge to individual users. Fees for reproduction and dissemination services may be waived by the NSSDC Director if:

(i) The cost of collecting the fee would be an unduly large proportion of the amount of the fee.

(ii) The data furnished are required to accomplish a research task approved by NASA Headquarters or field installations.

(iii) The data are to be used by a Federal, State, or local government agency or by nonprofit organization.

[FR Doc. 78-11862 Filed 5-1-78, 8:45 am]

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-89403		

4. Title and Subtitle	5. Report Date
Space Research Data Management in the National Aeronautics and Space Administration	November 1986
	6. Performing Organization Code P
7. Author(s) George H. Ludwig	8. Performing Organization Report No.
	10. Work Unit No.
9. Performing Organization Name and Address Office of the Chief Scientist	11. Contract or Grant No.
	13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Space-related scientific research has passed through a natural evolutionary process. The task of extracting the meaningful information from the raw data is highly involved and will require data processing capabilities that do not exist today. This report presents the results of a three-year examination of this subject, using an earlier report as a starting point. The report's general conclusion is that there are areas in which NASA's data management practices can be improved and recommends specific actions. These actions will enhance NASA's ability to extract more of the potential data and to capitalize on future opportunities.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Space research data information systems	Databases	Unclassified - Unlimited
Information technology	Repositories	
Data system planning	Active databases	
Data policy	Archives	
	Data centers	Subject Category 82

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	104	A06

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1986